[Inner endpaper]

HEC MONTRÉAL
École affiliée à l'Université de Montréal

Dual Perspective on Mental State Inferences in Human-Computer Interaction
A three-essay thesis

par
Théophile Demazure

Thèse présentée en vue de l'obtention du grade Ph. D. en administration
(option Technologies de l'information)

Juin 2023

HEC MONTRÉAL

École affiliée à l'Université de Montréal

Cette thèse intitulée :

Dual Perspective on Mental State Inferences in Human-Computer Interaction :
A three-essay thesis

Présentée par :

Théophile Demazure

a été évaluée par un jury composé des personnes suivantes :

Ana Ortiz de Guinea Lopez de Arana
HEC Montréal
Présidente rapporteuse

Pierre-Majorique Léger
HEC Montréal
Directeur de recherche

Marc Fredette
HEC Montréal
Codirecteur de recherche

Laurent Charlin
HEC Montréal
Membre du jury

Luc Cassivi
ESG UQAM
Membre du jury

Anthony Vance
Virginia Tech
Examinateur externe

Camille Grange
HEC Montréal
Représentante du directeur de HEC Montréal

Résumé

Cette thèse explore diverses méthodologies pour estimer l'état mental des utilisateurs pendant l'interaction homme-machine (IHM) avec les technologies de l’information (TI). Les processus mentaux jouent un rôle important dans l'utilisation des TI et peuvent être influencés par de multiples facteurs, y compris les caractéristiques de la technologie elle-même, les choix de conception et la tâche à accomplir. Par conséquent, la compréhension des états mentaux est cruciale dans l'étude des artefacts technologiques nécessitant une interaction humaine. Pour relever ce défi, la thèse adopte une approche NeuroIS, tirant parti des connaissances de la littérature en neurosciences cognitives pour faire progresser le développement et la mesure des états mentaux durant l’utilisation des TI.

Chaque chapitre de cette thèse contribue de manière unique à l'objectif global. Une revue de littérature est effectuée pour identifier les défis actuels de l'estimation de l'état mental en neurosciences appliquées en IHM. Cette recension des écrits a mis en évidence que ces défis découlent, en partie, d'une conceptualisation limitée des inférences psychophysiologiques et d'une adoption limitée des approches analytiques de pointe. Les chapitres suivants visent à remédier à ces limitations par le biais de deux méthodologies distinctes. La première approche utilise une méthodologie axée sur les données, en recourant à l'apprentissage automatique pour le décodage de l'état mental. L'objectif est de généraliser les réponses neurophysiologiques observées à partir de tâches synthétiques à une tâche naturelle. La seconde approche adopte une perspective hypothético-déductive et s'appuie sur la compréhension actuelle de l'intégration multisensorielle et de son interaction avec l'attention. En s'appuyant sur ce cadre théorique, une mesure de l'attention basée sur la perturbation des sens est développée et testée dans des tâches naturelles.

En adoptant ces méthodologies, cette thèse contribue à faire avancer le domaine des NeuroIS et la mesure de l'état mental au cours de l'utilisation des TI.

Mots clés : NeuroIS, HCI, apprentissage automatique, potentiel évoqué, électroencéphalographie, charge mentale, intégration multisensorielle, attention.
Méthodes de recherche : Revue de littérature, Expérience en laboratoire

Abstract

This thesis explores various methodologies for estimating mental state constructs during Human-Computer Interaction (HCI) in Information Systems (IS). Mental processes play a significant role in the use of information technology and can be influenced by multiple factors, including the characteristics of the technology itself, design choices, and the task at hand. Consequently, understanding mental states is crucial in studying technological artifacts requiring human interaction. To tackle this challenge, the thesis adopts a NeuroIS approach, leveraging cognitive neuroscience insights to advance the development and measurement of mental state constructs during IS use.

Each chapter of this thesis contributes uniquely to the overarching goal. A scoping review was conducted to identify current challenges in mental state estimation in applied neuroscience within HCI and IS. The review highlighted that these challenges arise, in part, from a limited conceptualization of psychophysiological inferences and a slow adoption of state-of-the-art analytical approaches. The subsequent chapters aim to address these limitations through two distinct methodologies. The first approach utilizes a data-driven methodology, employing advanced machine learning techniques for mental state decoding. The objective is to generalize the observed physical brain patterns from synthetic tasks to naturalistic settings. The second approach takes a more hypothetico-deductive perspective and draws on the current understanding of multisensory integration and its interplay with attention and various bottom-up and top-down factors. Building upon this theoretical foundation, a perturbation-based measure of oriented attention in EEG is developed and tested within naturalistic tasks specific to the IS domain.

By adopting these methodologies, this thesis contributes to advancing the field of NeuroIS and the measure of mental state over time during IS use.

Keywords: NeuroIS, HCI, machine learning, evoked-related potential, electroencephalography, mental workload, multisensory integration, attention
Research methods: Literature Review, Laboratory Experiment

Table of contents

List of tables

List of figures

À mes parents,

Acknowledgments

Je tiens à exprimer mes remerciements particuliers à Pierre-Majorique Léger et Marc Fredette pour leur soutien inconditionnel et leurs conseils précieux tout au long de mes études doctorales.

Un immense merci à mes parents pour leur soutien indéfectible tout au long de mes études. C'est l'aboutissement de nombreuses années de travail intensif, de sacrifices et de dévouement.

Je tiens également à témoigner ma reconnaissance envers Gilbert Babin, qui m'a offert les opportunités qui ont suscité mon intérêt pour l'enseignement, à Alexander Karran pour son mentorat et nos échanges stimulants, et à Sylvain Senecal et Constantinos Coursaris pour leurs conseils judicieux.

Une pensée spéciale est destinée à mes collègues et amis du doctorat Sara-Maude Poirier, Antoine Ouimet, Mario Passalacqua, Michael Dolezek, Zoubier Tkiouat et Burak Öz.

J'exprime aussi ma gratitude à tous ceux avec qui j'ai eu l'opportunité de travailler et collaboré, les étudiants à qui j'ai enseigné, mes collègues du doctorat et les professeurs qui m'ont accompagné tout au long de mon parcours. J'ai aujourd'hui l'honneur de pouvoir les compter comme collègues pour les années à venir.

Enfin, un remerciement au Fonds de recherche du Québec - Nature et technologies et à IVADO, qui m'ont octroyé une bourse d'excellence au doctorat, pour leur soutien financier qui a contribué à la réalisation de cette recherche. Ma gratitude s'étend également à CAE inc., D-BOX et la Chaire de recherche industrielle CRSNG - Prompt en expérience utilisateur pour l'appui financier et l'accès aux simulateurs qui ont été utilisés dans cette recherche.

Preface

This thesis, authored by Théophile Demazure, represents an original and unpublished research conducted under the guidance of Dr. Pierre-Majorique Léger and Dr. Marc Fredette.

Dr. Alexander Karran graciously participated in chapters 2 and 3. Yasmine Maurice actively participated in the search and screening process of chapter 2.

Chapter 1
Introduction

1.1 Thesis Motivation

This thesis aims to advance measurement methods for estimating mental state constructs during human-computer interaction (HCI) in Information Systems (IS). Mental processes (e.g., cognitive load, engagement, immersion) are omnipresent during the use of information technology. They can be influenced by multiple artifact characteristics (e.g., technology itself, design choices, or the task at hand), making it an important preoccupation in the study of technological artifacts at the individual level. In order to address it, NeuroIS pushes for the use of neurophysiological tools to study the neural substrates of emotional, cognitive, and social processes (Dimoka et al., 2012; Loos et al., 2010; vom Brocke et al., 2020). This thesis engages with this call for research by drawing on cognitive neuroscience literature to contribute to mental state constructs developments and measurements in IS.

Neurophysiological methods offer the capability to uncover users' automatic and unconscious mental states during computerized tasks (Dimoka et al., 2011; Loos et al., 2010). These techniques enable a rich measurement of users' mental states, which in turn can offer new insights into user cognition during interactions with systems (Riedl et al., 2014; vom Brocke et al., 2020). By measuring automatic brain processes, we can enhance the measurement and deepen our understanding of information systems constructs (De Guinea et al., 2013; Tams et al., 2014).

In the recent decade, information system researchers have explored and utilized various methods, tools, and neuroscientific measurements. A review by Riedl et al. (2020) examined 73 articles from the past basket of eight information systems journals, revealing a substantial body of literature encompassing measures for both the automatic nervous system (ANS) and central nervous system (CNS). The study highlighted the relative frequency of various neuroscientific instruments used in empirical research, with electroencephalography (CNS), heart rate (ANS), skin conductance (ANS), functional magnetic resonance imaging (CNS), eye-tracking (pupil dilatation) (ANS), and facial electromyography (ANS) being the most commonly employed. IS researchers applied these instruments to infer constructs related to emotions, stress, attention, or trust.

This thesis focuses on electroencephalography (EEG), the most widely utilized instrument in IS (Riedl et al., 2020). This instrument records the electrical activity resulting from the activation of large numbers of neurons at the surface of a human scalp (see box 1.). According to Müller-Putz et al. (2015), EEG's popularity stems from its relatively low cost compared to other neural imaging techniques like fMRI. The relatively affordable acquisition price and cost per subject make it accessible for researchers. Additionally, EEG offers excellent temporal resolution, typically ranging from approximately 250 to 1000 Hz, and it is relatively non-intrusive for HCI studies. However, its spatial resolution is significantly limited, approximately 1 cm compared to fMRI. However, EEG signals are susceptible to environmental and internal interferences, such as body movements and eye blinks. Nonetheless, EEG is well-suited for computerized tasks, such as those in HCI, as they often involve computer usage, a seated position, and minimal body movement. In this context, EEG enables the investigation of automatic and unconscious cognitive mechanisms in the central nervous system without disrupting the ecological validity of IS tasks.

Box 1

Electroencephalography (EEG) foundations from Müller-Putz et al. (2015)

The ***neurons*** are the fundamental elements of the nervous system. They communicate with each other through chemical and electrical signals. A neuron is a nerve cell that receives, processes and sends information. Information is conveyed when neurons receive electrical stimulation from other neurons and release the chemical substance to the synapse.

The ***electroencephalogram*** records the electrical activity of neurons within the cerebral cortex via electrodes. Electrodes are placed at the scalp's surface and record the electrical signal generated by the synchronous excitation of neurons. Because neurons are microscopic and electrodes cover a small portion of the scalp, a large population of neurons in an outward direction must activate altogether to detect the sensors. This measurable change is the result of the summation of post-synaptic potentials that occur in clusters of neurons at the same time. EEG activity is relatively small, measured in microvolts (mV).

Event-related potentials (ERP) are physical brain responses originating from external stimuli and reflecting sensory, cognitive, or motor events. ERP signal is small (~ 5-20 mV), and it is difficult to decipher it from the other physical processes in the brain. An analytical technique is to record the same time-locked stimulation multiple times and average the electrical response. This procedure increases the signal-to-noise ratio.

Neural Oscillations are the rhythmic and repetitive electrical activity generated spontaneously or in response to external stimulation by clusters of neurons. The electrical signal contains multiple concurrent oscillations from different frequencies. To uncover them, the electrical signal is decomposed into distinct frequency bands (e.g., delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz)) with mathematical processes such as Fourier analysis.

Methodological contributions are emerging in NeuroIS for neurophysiological tools (Léger et al., 2014; Riedl et al., 2014; vom Brocke & Liang, 2014), but ones specific to instruments and methods to relate them to psychological constructs are still needed. Müller-Putz et al. (2015) proposed a first methodological introduction to EEG destined for IS researchers, but more research should focus on how to use those methods and incorporate in our practice efficiently. As noted by Riedl et al. (2014), "more research contributing to the systematic development of a NeuroIS research methodology is needed" (p. ii). This thesis directly engages in this call and transfers knowledge and techniques from cognitive neuroscience to NeuroIS.

However, adopting methodologies from neuroscience to applied contexts, such as in IS, is not effortless and comes with rigorous experimental constraints. By methodology, we refer to Mingers (2001) definition: the "structured set of guidelines or activities to assist in generating valid and reliable research results" (p. 242). It comprises many techniques or methods that need to be used when necessary, and neuroscience has some specific ones that may interfere and force compromises in IS research. Moreover, IS papers using these methodologies should be assessed based on the standard rigors expected in the related field of neuroscience (i.e., cognitive neuroscience and EEG) (vom Brocke & Liang, 2014).

Electroencephalography methods bear important considerations and limitations when transposed to less controlled fields such as HCI research. Most of these limitations are due to three elements and their interactions: (1) the neurophysiological measurement instrument itself, (2) the functioning of the brain, and (3) the environment's effect on those two. These elements are highly contextual and must be addressed in a particular study.

To cite a few, EEG measurements are prone to noises. It is even impossible to record signals without contamination (Müller-Putz et al., 2015). Physiological noises come from the participant itself and are omnipresent. It includes heartbeats, muscle movements, ocular noises (i.e., blinks, vertical and horizontal eyeball movements), and ongoing parallel brain processes. Environmental noises can come from external electrical outlets, subject grounding, or electrode contacts. These noises are more prevalent in naturalistic and rich tasks such as in HCI research.

Moreover, physical brain activity has many-to-many relationships with constructs (Cacioppo et al., 2007) and concurrent neurophysiological processes (e.g., working memory interacts with attention). Thus, it requires parsimonious experimental designs that carefully manipulate the targeted phenomena or constructs. It implies a good conceptualization of the IS task (e.g., characteristic of interaction, cognitive demand, top-down factor, stimuli saliency). Features of the task generate specific brain responses. Manipulating these features without considering the neural substrates might lead to uncontrolled change and the impossibility of explaining what drove it. This last point is particularly challenging in IS; such a task is often complex and involves intertwined brain functions.

To tackle those challenges, significant advances for mental state measurement during complex interactions are being made in the field of neuroscience and more applied fields such as neuroergonomics (Frey et al., 2018; Johnson & Proctor, 2013; Lotte & Roy, 2019; Mehta & Parasuraman, 2013) or brain-computer interfaces (BCI) (Zander & Kothe, 2011). In parallel, recent developments in machine learning techniques combined with domain-specific methodologies are increasingly employed (Alzahab et al., 2021; Craik et al., 2019; Roy, Hubert, et al., 2019). Unfortunately, NeuroIS still lags in developing and leveraging robust and reliable state-of-the-art mental state measurement, enabling adequate ecological validity, especially during prolonged complex interactions. This need has been identified as two core NeuroIS opportunities, i.e. (1) the capture of hidden mental processes that are challenging to measure with traditional methods and (2) mental state measures to directly inform the usage or the design outcomes of technological artifacts (Dimoka et al., 2011; vom Brocke et al., 2020).

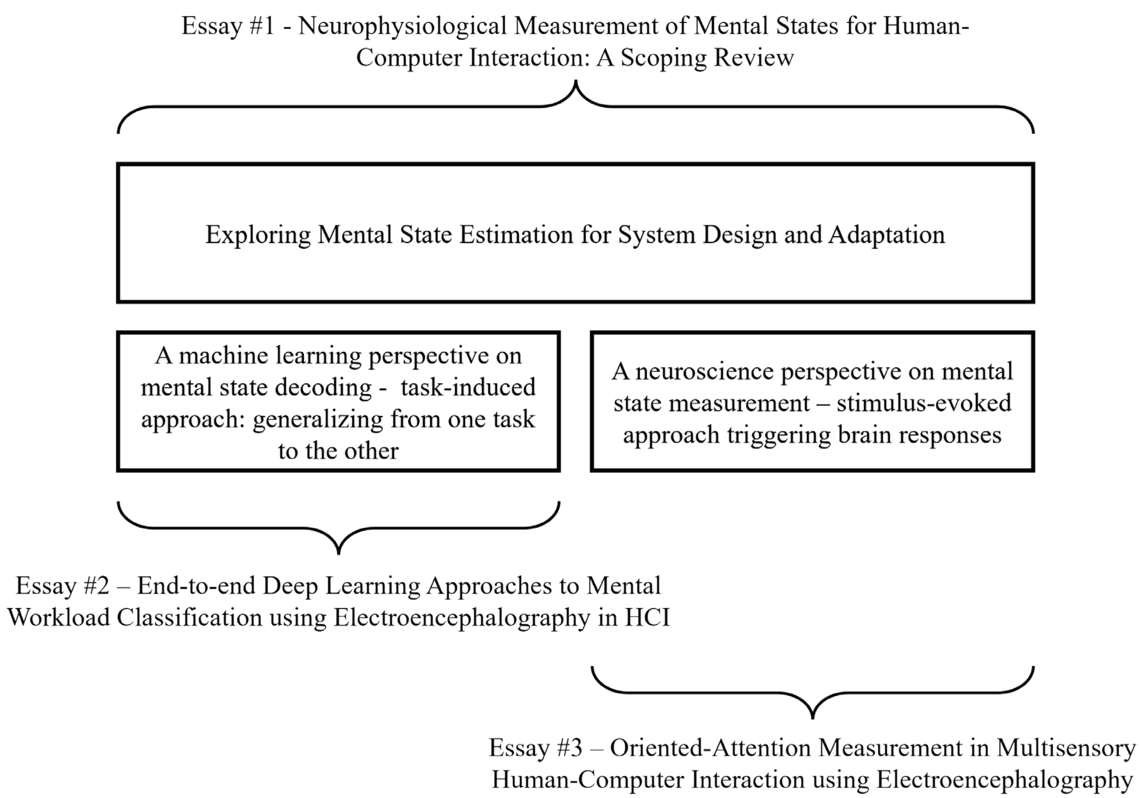
This thesis aims to enhance methods for estimating mental states by integrating research from HCI with neuroscience. The objective is to provide a conceptual framework that bridges methodological perspectives with realistic application settings. These methodologies are applied to ecologically valid HCI tasks. Figure 1 visually depicts how the three essays interlock.

The thesis is structured as follows. The first essay (Chapter 2) comprises a scoping review that surveys the literature on measuring mental states in applied fields of neuroscience. In chapters 2 and 3, we conducted empirical tests on two potential techniques for inferring mental states in simulator-based tasks under controlled experimental conditions. Chapter 3 specifically focuses on working memory and utilizes state-of-the-art end-to-end deep learning techniques to decode the mental state. Chapter 4 delves into multisensory integration and explores its interaction with attention to gain a better understanding of its role in complex human-computer interactions.

Overall, the thesis embraces the co-evolutionary perspective of IS and Neuroscience research, which will be outlined in the subsequent sections. Through the application of distinctly different mental state inference techniques in these two chapters, we hope to foster an insightful conversation on the challenges encountered in achieving our objective of robust mental state inferences.

Figure 1

Thesis framework



Before engaging in the contents of the thesis, it is crucial to explain the concepts used throughout the manuscript. Subsequently, the rationale behind our principal study settings, simulator-based environments, will be discussed. Such tasks offer unequivocal advantages in striking a balance between ecological and internal validity, essential for robust inferences regarding mental states in NeuroIS. Lastly, we will summarize the various essays in this thesis and demonstrate how they align with the adopted perspective.

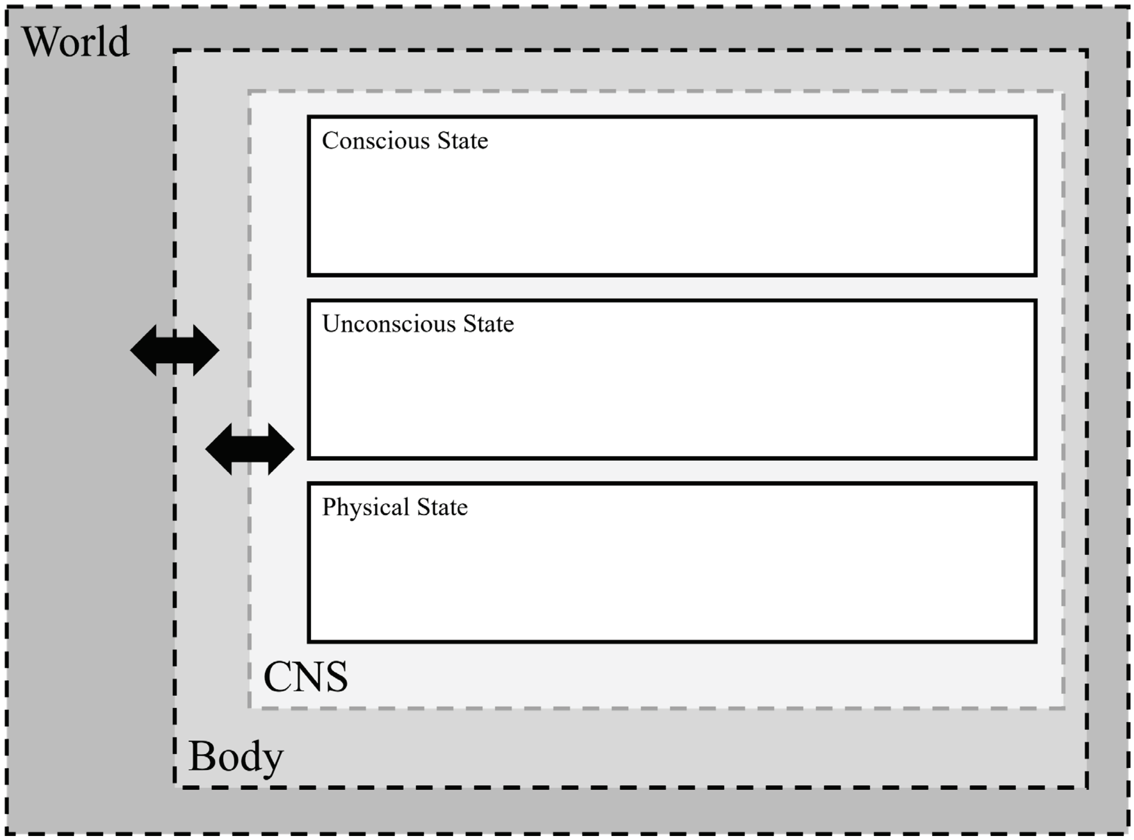
1.2 Concepts Clarification and Definitions

This thesis clearly distinguishes between mental states and the underlying brain processes. It is essential to recognize that nothing in the brain or within an individual mind is fixed or static. The physical processes occurring in the brain involve ongoing neural activities that can spontaneously arise or be triggered in response to stimulation. This dynamic nature is also observed in conscious sensory states, such as the experience of touch or pain, as well as in attitudes, such as beliefs and thoughts. Additionally, conscious cognitive states (e.g., attention, workload) and unconscious ones (e.g., sensory integration, working memory) exhibit inherent instability. Although these "states" exist at different levels of abstraction, they all refer to concepts that possess a certain degree of constancy, even though the underlying phenomena themselves may be dynamic. Therefore, the terms "process" and "state" are used interchangeably in describing relatively stable events occurring within the individual and the brain.

In the previous paragraph, we just mentioned many concepts without introducing them. We might have guided an intuition by carefully choosing adjectives relating to "states." However, definitions are still necessary to avoid confusion; "Mental states" is a broad concept. This thesis distinguishes three conceptual levels when referring to mental states: physical, cognitive, and conscious (see Figure 2).

Figure 2

The three conceptual levels of mental state



Conscious states are phenomenal states, the continually changing subjective and experiential aspects of the mind. A phenomenal state can be defined as a “mental state that is individuated by what it is like for one to be in it” (Chudnoff, 2015). Thus, a state is conscious if the ones experiencing it are, to some extent, aware of experiencing it (Rosenthal, 1986). It comprises intentional states like beliefs, thoughts, desires, attitudes, or judgments. It also relates to affective experiences such as emotions or moods, bodily experiences like pain or hunger, and cognitive experiences like boredom, immersion, and concentration (Schwarz & Clore, 2007). Some phenomenal states span across multiple of them, like stress can be felt via the body physically and physiologically, but also emotionally and cognitively.

The *unconscious states* refer to the unconscious and automatic processes happening within the brain during a task. We can all distinguish conscious mental processes we are aware of (e.g., thoughts, feelings, reasoning processes, pain). However, while we experience these processes, we are unaware of the brain's underlying cognitive mechanisms and physical processes from which they emerge. Users are not mindful of their existence (De Guinea & Markus, 2009; De Guinea et al., 2013). They are still theoretical constructs that are sub-components of conscious processes while they are becoming experiential.

The *physical states* correspond to the measurable physiological responses of the body. These states often reflect cognitive mechanisms associated with both conscious and unconscious activities. For instance, they encompass the electrical activity measured by EEG, oxygenation levels assessed through near-infrared spectroscopy, or the magnetic field generated by electrical currents measured by magnetoencephalography. Neurophysiological instruments and methods play a crucial role in capturing these physical patterns associated with underlying neurophysiological processes.

However, as we aim for ecological validity and the naturalness of context and tasks in IS, it becomes crucial to consider cognition and mental states as embodied. Embodied cognition recognizes that mental states are not solely dependent on the brain, but also on the body and the surrounding *world* (Northoff, 2018). This perspective contrasts with the experimental approach traditionally adopted in cognitive neuroscience, which often studies cognitive processes in isolation (Ladouce et al., 2016; Stangl et al., 2023). Viewing the body as a simple input-output system where sensory integration and decoding only influence behaviors is an incomplete understanding. It generally relies on bottom-up influences on cognition. Instead, embodied cognition emphasizes the bidirectional influences between cognition and the body within its environment.

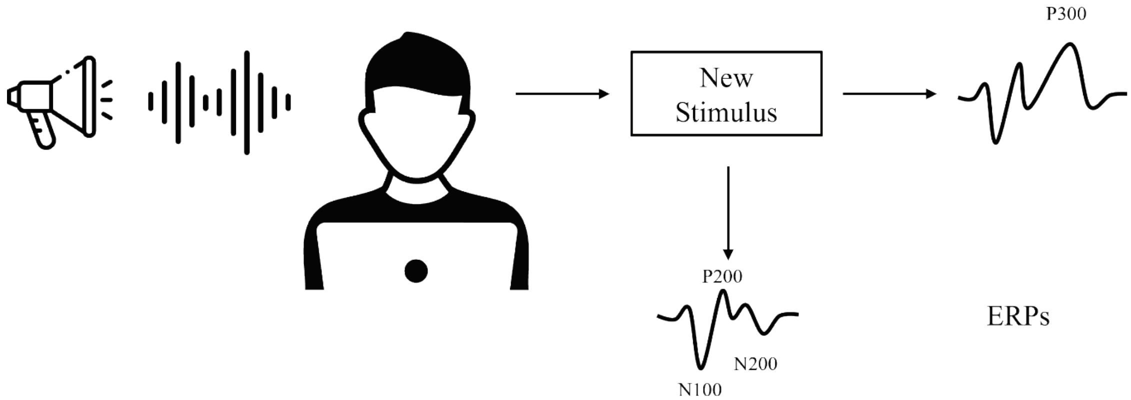
Numerous pieces of evidence highlight the influence of top-down executive control factors on cognition along with bottom-up influences (Ladouce et al., 2016, 2019; Macaluso et al., 2016; Talsma et al., 2010). This suggests that cognition is not influenced only by sensory input but also by higher-level cognitive processes (also called executing control, top-down factors) that shape our perception and behavior. Therefore, conceptualizing the user's mental states within the context of the task's interaction and the environment is essential for understanding cognition during IS tasks. This approach is particularly emphasized in Chapter 4 - Essay #3, where we delve into the topic more deeply.

As a concrete illustration of those concepts, we can consider the well-known context-updating theory of sensory input, which explains the cognitive process of receiving auditory stimulation (Figure 3). When an auditory stimulus is encountered, it starts a complex cognitive process until its conscious detection. The auditory stimulation is received, encoded, and enters the processing system. Resources are allocated to attend to and process this auditory input. Within the working memory, the internal representations of the stimulus are compared with existing representations and updated if necessary (Polich, 2007).

From a neurophysiological perspective, the various components of this process can be directly associated with observable physical states. The neural responses to auditory stimulation can be measured and analyzed using event-related potentials (ERPs). Within the context of ERPs, the early responses (such as N100, P200, and N200) observed in the waveform correspond to the initial stages of auditory processing. These early neural responses reflect the early analysis of the sensory evoked potentials of the auditory stimulus at the physical level. The late positive response (P300) may signify the updating of the new stimulus representation. This late physical response is absent when no change is detected. It is important to acknowledge that this is a concise and simplified description of the sensory integration of auditory stimuli. For a more comprehensive perspective, refer to Gonsalvez and Polich (2002), Polich (1989, 1998), Polich (2007), Polich and Heine (1996), or Polich and Kok (1995). It should be noted that both top-down and bottom-up factors can influence this entire cognitive process; Chapter 4 extensively explores this subject. Nonetheless, these processes illustrate the relationship between physical, unconscious, and conscious states.

Figure 3

Context-updating theory of sensory input



Note. Figure adapted from Polich, 2007

In this particular example, cognitive mechanisms manifest themselves as physical responses elicited by a discrete perturbation, namely, an auditory stimulus. In this case, mental states represent transient equilibria that persist for a few milliseconds, as evidenced by event-related potentials (ERPs). However, changes in mental states can also occur over more extended periods of time, giving rise to recurring patterns of physical responses, such as oscillations. To account for both conditions, we adopt two perspectives on mental states at the physical level: task-induced and stimulus-evoked. The *task-induced perspective* refers to the brain activity elicited by the task being performed. It manifests itself as an ongoing electrical activity that is influenced by the characteristics of the task, modulated by internal user factors and is often studied in a continuous manner.

On the other hand, the *stimulus-evoked perspective* focuses on changes in brain activity induced by specific external stimuli (Müller-Putz et al., 2015). These stimuli are typically presented to the subject at a specific time and within a defined sensory modality, such as visual, auditory, or vestibular. The stimuli can be related or unrelated to the task at hand and are operationalized using time-locked paradigms. Task-induced and stimulus-evoked perspectives are closely related to the psychological constructs that can be studied and their operationalization within the research context.

Task-induced and stimulus-evoked brain activities are to be contrasted with spontaneous activity or resting state concepts. *Spontaneous activity* is the “neural activity that is generated within the brain itself independent of any external stimuli from outside the brain, including interoceptive stimuli from the body and exteroceptive stimuli from the world” (Northoff, 2018). It refers to the ongoing physical activities in the brain present even in the absence of any stimulation.

Resting state often alludes to the operationalization to measure spontaneous activity and its behavioral condition (Northoff, 2018). During this period, no sensory systems are stimulated. It is often implemented as eyes opened or closed conditions with a fixed and stable fixation during a defined time.

In short, task-induced and stimulus-evoked activity is operationally generated and tested with a particular task/event, while spontaneous activity goes on perpetually (Müller-Putz et al., 2015). By contrast, within an individual, the elicited physical neural response can be the difference between the two. However, it is now known that the separation of spontaneous and task-related brain activity is much more nuanced as they might modulate each other (Northoff, 2018). However, their relationships are beyond the scope of this thesis and won’t be addressed further.

To put this thesis in perspective from this clarification, Chapter 3 - Essay #2 leverages a task-induced paradigm for mental state inference in fully naturalistic real-world research in a simulator out of the laboratory. Chapter 4 - Essay #3 uses a stimulus-evoked paradigm for mental state measurement in a naturalistic laboratory experiment.

1.3 Co-evolutionary perspective of IS and Cognitive Neuroscience

This thesis has been profoundly influenced by Churchland (1989) and Northoff (2018) books on the mind-brain problem. One of the main propositions of Churchland (1989) in her seminal books is that ‘[...] the theoretical framework resulting from the co-evolution of neuroscience and psychology is bound to be superior to folk psychology [...]’. This thesis embraces a similar co-evolutionary perspective between IS and cognitive neuroscience and within IS. Here, we strive to integrate perspectives from cognitive neuroscience, IS, and NeuroIS, contributing to each of these disciplines. Figure 4 depicts the relationships between the different fields and their relevance paths.

1.3.1 "Within" relevance path: motivate and enrich

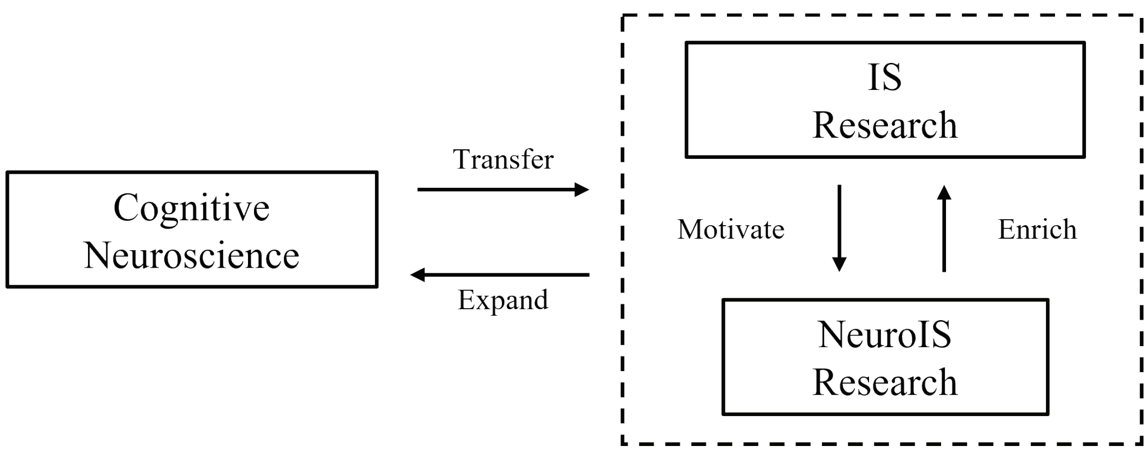
The "within" relevance path in IS research motivates the selection of constructs and phenomena to which NeuroIS can contribute. In return, NeuroIS enriches understanding of conscious states (i.e., constructs) with related unconscious states and their measurements at the physical level in IS tasks. In this perspective, the IS and NeuroIS viewpoints are co-dependent. IS elucidates perceptual and intentional states at the conscious level, while NeuroIS explores the underlying unconscious cognitive mechanisms and their corresponding brain physical processes. As Kirwan et al. (2023) highlighted, a significant motivation in IS research is to unveil the "black box" of the brain, which can be achieved through two approaches. The first approach involves employing neurophysiological methods to understand the brain's cognitive processes better. The second approach entails utilizing behavioral methods to investigate how inputs and outputs manifest in the user. By employing these approaches, researchers aim to describe the ongoing mechanisms and enhance our comprehension of cognitive processes and behaviors during naturalistic IS tasks.

In this thesis, the chosen approach entails identifying pertinent psychological constructs related to mental states and deconstructing their dimensions into lower-level constructs at the cognitive mechanisms level. Subsequently, neurophysiological measurements are associated with these constructs. For instance, Chapter 3 - Essay #4 concentrates on a particular dimension of mental workload, namely working memory. Chapter 4 - Essay #4 focuses on the unconscious process of multisensory integration and attentional orientation. Understanding and quantifying this mechanism is crucial for investigating the underlying attentional mechanisms associated with conscious constructs like immersion, presence, and concentration.

The "within" relevance path contributes to the IS field in several significant ways. Firstly, it enhances our understanding of the cognitive mechanisms underlying key conscious states within IS. Providing instruments and methodologies for measuring these cognitive mechanisms at the physical level opens up new avenues for assessing IS constructs (Dimoka et al., 2012) and capturing their domain of unconscious processes (Tams et al., 2014). Thus, low-level explanations of the brain processes might increase our understanding of IS theories and constructs and even maybe revisit them. Additionally, within the context of Design Science Research (DSR), neurophysiological measurements can inform the design of IT artifacts (Riedl & Léger, 2016; vom Brocke et al., 2020). They can be utilized to investigate the impact of technology on users' cognitive mechanisms during usage and uncover differences arising from various design choices.

Figure 4

Co-evolutionary Perspective of IS, NeuroIS and Cognitive Neuroscience and their Relevance Paths



1.3.2 "Between" relevant path: transfer and expand

The "between" relevance path describes the transfer of descriptive and prescriptive knowledge from neuroscience to IS. While neuroscience provides insight into the brain, methods, and instruments, enabling contributions to IS, IS, in turn, extends cognitive neuroscience knowledge by generalizing hypotheses to (quasi-)naturalistic settings and tasks. This view emphasizes the two-way nature of the relevance path between IS and cognitive neuroscience. Cognitive neuroscience's understanding of the brain, methodologies, and instruments is initially transferred to IS, enabling the aforementioned contributions. However, IS also has the potential to expand the existing knowledge in cognitive neuroscience. Generalizing the understanding of the brain to naturalistic environments and tasks is a significant objective pursued by neuroscience researchers who aim to comprehend human cognition in real-world scenarios (Matusz et al., 2019; Northoff, 2018; Stangl et al., 2023). By informing NeuroIS research with the current domain of knowledge derived from laboratory experiments, a step towards real-world studies in (quasi)naturalistic computerized tasks is taken, which expands the ecologically valid understanding of cognition. Therefore, NeuroIS contributes to the field of cognitive neuroscience in this way.

In this thesis, the different chapters contribute to the "between" relevance path in different ways. Chapter 3 - Essay #2 focuses on improving mental state estimation in EEG using a data-driven approach and state-of-the-art machine learning techniques, specifically end-to-end deep learning architectures. Leveraging these advanced methodologies transfers knowledge and methodological techniques from cognitive neuroscience and machine learning to improve mental state estimation. On the other hand, Chapter 4 - Essay #3 builds on a conceptual framework derived from cognitive neuroscience that explores multisensory integration and its relationship to attention. This chapter also contributes to the "between" relevance path by transferring knowledge from cognitive neuroscience to the field of IS. It also provides novel evidence for applying this framework in naturalistic tasks, shedding light on how multisensory integration and attention operate in real-world settings. Both chapters exemplify the dissertation's commitment to transferring knowledge, methods, and frameworks from cognitive neuroscience to the IS field while demonstrating their practical application in naturalistic tasks.

In conclusion, while IS can benefit from a better understanding of neural mechanisms during interactions, it is essential to note that IS theories can and will continue to advance without explicitly referencing the brain processes underlying its constructs. Nevertheless, NeuroIS and cognitive neuroscience can contribute significantly to enriching the field's current understanding at the individual level. At the same time, IS tasks are inherently naturalistic. Therefore, the transfer and validation of hypotheses from cognitive neuroscience to NeuroIS can contribute to the goal of cognitive neuroscience by validating hypotheses in naturalistic settings. By bridging neuroscience and IS, we will gain insight into the brain processes involved in commonly used mental state constructs and foster the co-evolution of theoretical frameworks, further improving our understanding of the brain and IS constructs in tandem.

A note of caution, we should address some pitfalls/limitations of the approaches we will expand on below and in light of this section. First, we selected psychological constructs (conscious state) as a starting point for both empirical studies in this thesis and broke them down into relevant cognitive mechanisms that we can operationalize with neuroscience methods. The contrary can be done and might lead to interesting new constructs. Secondly, breaking constructs to attain testable concepts in neuroscience can lead to the loss of meaning of the starting construct. Finally, the operational definition of the construct at the physical brain level can give a false impression that a single mental process has been addressed and manipulated (Kotchoubey et al., 2016). Brain processes are interdependent and involve many simultaneous components. The number of mental states and variations is infinite (Haynes & Rees, 2006), while the manipulation or labels are limited.

1.4 Simulator-based Task for NeuroIS

NeuroIS often addresses real-world, complex, and dynamic phenomena. However, cognitive neuroscience and neurophysiological tools are challenging to use effectively in those environments. Indeed, many brain processes have been established in cognitive neuroscience with controlled, simplified, and artificial stimuli (Matusz et al., 2019). These paradigms put particular care into creating experimental paradigms manipulating well-defined brain functions. One core advantage of using artificial stimuli is their parametrization (Felsen & Dan, 2005). Varying stimuli parameters (e.g., orientation, spatial position, tones) facilitate the study of fine-grained changes in neural responses. Unfortunately, these findings utilizing these paradigms do not generalize well to more complex and natural tasks (Felsen & Dan, 2005; Stangl et al., 2023). This limitation is not restricted to brain states but also behaviors (Ladouce et al., 2016), artificial stimuli produce different behaviors in control contexts than in-situ.

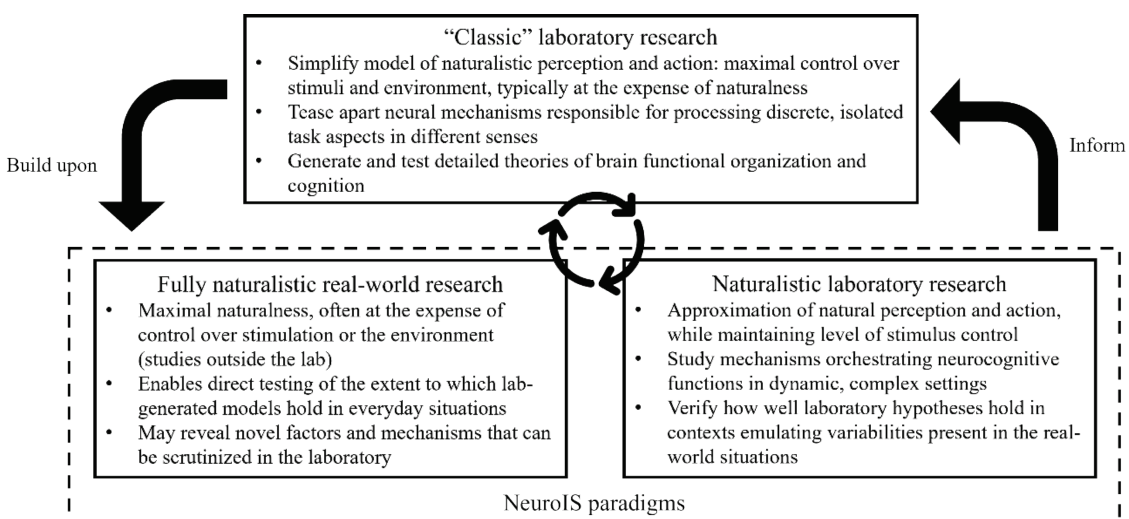
However, cognitive neuroscience in naturalistic experiments is considered to be feasible and even encouraged (Matusz et al., 2019; Stangl et al., 2023). Authors have shown that some physical responses can be reliable even during an uncontrolled task like watching a movie (Hasson et al., 2010; Spiers & Maguire, 2007). Hasson et al. (2010) demonstrated in a review the existence of shared cortical responses across individuals during free watching. They also showed that robust brain responses were found within individuals or sub-groups. Ladouce et al. (2019) revealed that neural correlates of attention are reduced in real-world and natural behavior like walking in a dynamic environment. Thus, attentional processes and physical substrates in the brain are measurable with EEG and robust to real-world settings.

Those findings might seem insignificant for NeuroIS, but it is quite the contrary. It reveals that applying cognitive neuroscience methods to IS phenomena is feasible despite the relaxation of the experimental internal validity. Furthermore, the quest for finding robust brain responses within and across users and their relation to psychological constructs might not be in vain. In the context of cognitive and social neuroscience, Matusz et al. (2019) propose that naturalistic research should build upon the detailed and controlled hypothesis tested at the expense of external validity (Figure 5). Thus, naturalistic laboratory studies can imitate real-world stimulation while keeping a relative stimulus and environmental control level. It offers a good trade-off between external validity and internal validity. Likewise, fully naturalistic real-world research allows building upon classic and naturalistic laboratory research. In those settings, stimulation and environmental control are impossible. However, context specific paradigms and analytical techniques might permit the use of neurophysiological measurements as well (Bevilacqua et al., 2019; Ladouce et al., 2016, 2019; Matusz et al., 2019; Midha et al., 2021). Moreover, Stangl et al. (2023) ascertain that technological advances and preprocessing techniques advances reduce some of the limitations that naturalistic tasks impose on data quality. Studying the neurophysiological mechanisms of cognition and behavior in real-world HCI seems, as time passes and research goes on, more and more plausible and promising.

If we had to place NeuroIS experimental research on this schema, it would comprise quasi-naturalistic laboratory research to fully naturalistic real-world research for most studies. Even the most simplistic and artificial IS stimuli or artifacts are far from the level of control found in cognitive neuroscience. This limits the generalizability of the underlying understanding of the neurophysiological mechanisms and their physical response from synthetic tasks to HCI tasks. Thus, it is essential to use rigorous and aligned research paradigms that allow the creation of plausible inferences with knowledge of the limitations that our phenomena of interest bring.

Figure 5

Situating NeuroIS research paradigms in contrast to cognitive neuroscience opportunities and challenges of naturalistic research



Note. Figure adapted from Matusz et al., 2019.

Now, we are left with the initial challenge. How do we deal with ecologically valid artifacts/tasks and the minimal control level necessary to study mental states in NeuroIS? There are many conceptual and methodological solutions to explore, but we adopted the context of simulator-based tasks in this thesis. Simulated environments and tasks, also referred as microworld, have been extensively used in neuroergonomics, BCI, and HCI to study mental states (Brehmer & Dörner, 1993; Mehta & Parasuraman, 2013; Pope, 1995). Its use has even been proposed in NeuroIS (Loos et al., 2010). Similarly, augmented reality and virtual reality have been leveraged to study cognition to simulate naturalistic settings in laboratory experiments (Stangl et al., 2023). Recently, even the metaverse has been proposed to study human behavior (Gómez-Zará et al., 2023).

All these technologies share similar promises and advantages in experimental settings. They are adaptable technologies that can be used for research, interface design, and training. Simulators replicate real-world dynamical environments and offer a fine compromise between internal and external validity for quasi-real-world tasks in naturalistic settings (Brehmer & Dörner, 1993). Simulation-based tasks present significant advantages in research when studying mental processes—advantages that are vital for a rigorous study of mental states at the neurophysiological level in NeuroIS. These advantages are controllability, reproducibility, standardization, ease of comprehensive data collection, and safety.

Controllability allows the creation of parametric tasks that facilitate the manipulation of specific aspects within the simulator while keeping other parameters constant (De Winter et al., 2012). Controllability is necessary to manipulate IS artifacts and task properties and observe robust and reliable brain patterns. The implications are more than conceptual. Operationally, it also allows the combining of neurophysiological signals within similar conditions and contrasts them between conditions. Repeatability (Bland & Altman, 1986) is deeply linked to controllability and is necessary for NeuroIS when using neurophysiological tools. Repeatability, or test-retest reliability, refers to the variation in repeated measurements taken on a subject with the same instrument in the same condition (Bland & Altman, 1986). A good paradigm should ensure that the repetition of a task (i.e., the condition with specific parameters) will elicit similar brain responses and enable repeated and reliable measurement (Riedl et al., 2014). When using electroencephalography, the objective is to increase the signal-to-noise ratio (see Chapter 8 in Luck (2005, p. 261) for a detailed discussion). Very briefly, averaging the electrical signal will reduce the presence of noise or irrelevant brain processes that are not provoked by the task (Müller-Putz et al., 2015). Simulator-based tasks increase control over realistic tasks (Baldwin, 2019) and increase measurement reliability in NeuroIS.

Reproducibility and *standardization*, Simulator-based tasks can be replicated in another location or even for complementary findings within multi-study research (De Winter et al., 2012). Simulators are digital artifacts that can be easily copied and installed in another environment. Simulator-based tasks can be standardized to reduce real-world environments' randomness and uncontrolled nature. Reproducibility is an essential component of science, and even rigorous research can be subject to the issue. In a meta-analysis conducted on 100 seminal studies in psychology, researchers found that out of 97 that presented significant results, only 35 were replicable (Collaboration, 2015). This challenge is pervasive across many domains (Baker, 2016). Discussing those findings, Conrad and Bailey (2020) noted that NeuroIS conformed to the standard imposed in Neuroimaging. Unfortunately, Poldrack et al. (2017) showed that neuroimaging might not remain untouched by the problem. Nevertheless, reproducibility is an important feature of rigorous research, and simulator-based tasks can facilitate replications in NeuroIS.

Secondly, De Winter et al. (2012) noted that simulator-based tasks *ease the data collection* for behavioral and performance measurements, especially in a controlled simulated environment with standardized tasks. Moreover, synchronizing events and behavioral data within the simulator with concurrent neurophysiological signals is a significant advantage. It provides an accurate temporal representation of physical processes within the brain during the task (Ladouce et al., 2016). Moreover, researchers have already developed protocols to unify and synchronize time-series measurement with milliseconds precision, such as Lab Streaming Layer[1].

Thirdly, simulator-based tasks reduce *ethical* and *safety* concerns. It can be used to create unpredictable or unique conditions without risking dangerous situations for the subject (De Winter et al., 2012). IS tasks are often work-related, and interfering with someone's work can have unwanted consequences for the subject.

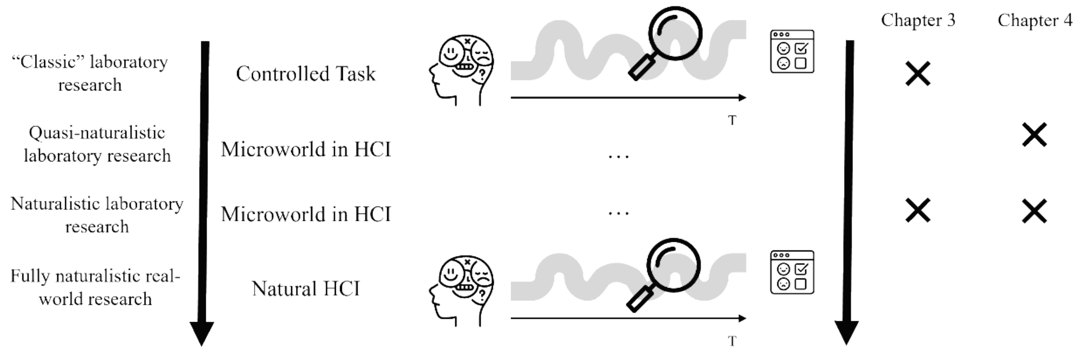
Naturally, simulator-based tasks possess some limitations. Firstly, low-fidelity simulators can lead to unwanted consequences (De Winter et al., 2012). Physical, perceptual, and behavioral fidelity negatively impact users' attitudes and behaviors. Fidelity can also impact the generalizability of the study. Secondly, they cannot encompass all the environmental conditions present in natural settings (Baldwin, 2019). Third, there is still limited evidence that compares real-world and simulator-based tasks (De Winter et al., 2012), even if some study shows similitude in behaviors between simulated and real-world tasks (Wang et al., 2010). Fourth, simulator-based tasks require an adaptation period (Baldwin, 2019). Therefore, users should familiarize themselves with the simulator controls and characteristics.

Researchers in IS developed microworld and argued for using simulation (Léger, 2006; Léger et al., 2011; Loos et al., 2010), but usage is still scarce. As an example of research using simulation in NeuroIS, Demazure et al. (2021) and Karran et al. (2018) used ERPsim (Léger et al., 2007) for the development of a neuro-adaptive system integrated to a business environment and task. Considering our previous arguments, simulation-based tasks can offer the experimental contexts that IS needs. The clear balance between control and naturalism it offers fits the identity of the research field. As IS relevant simulations are still limited, we selected "extreme cases" of realistic simulators to push the boundaries of feasibility. Generalisability to IS tasks might. Engaging with this perspective, we leverage simulators as our main experimental context. We believe that methodological learning can be generalized to more common IS situations.

[1] https://labstreaminglayer.readthedocs.io/

Figure 6

Thesis positioning in terms of experimental naturalness



In this thesis, we attempt to find cognitive mechanisms and measurement methods that generalize from synthetic and controlled tasks to naturalistic ones. We use simulation-based tasks to study mental states and measurement techniques: (1) working memory aspect of the mental workload in a flight simulator (Chapter 3 - Essay #2) and (2) the orientation of attention in a driving simulator (Chapter 4 - Essay #3). Figure 6 shows how both empirical manuscripts map in terms of the naturalness of the experiment. Both manuscripts address this challenge in different manners. Chapter 3 attempts to find recurring physical patterns between a synthetic and empirically supported experimental paradigm and a natural one. Chapter 4 manipulates the level of naturalness of the task to evaluate the generalizability of a physical response to neurophysiological mechanisms.

1.5 Application of the Concepts in the Essays

The perspective presented in the introductory discussion profoundly shaped the manuscripts comprising the present thesis. Table 1 provides an overview of how the concepts outlined in the introduction are implemented and applied in the research conducted throughout the thesis. In addition, it demonstrates the operationalization of these concepts within the research endeavors undertaken. The subsequent section presents a summary of each essay that forms the foundation of this thesis. Each essay begins with a brief introduction, followed by a description of its objectives and methodology. Finally, the contributions of each essay are outlined.

Table 1

Application of the Concepts in the two empirical essays

	Chapter #3 – Essay #2	Chapter #4 – Essay #3
Nature of the manuscripts	Empirical study, laboratory experiment	Empirical studies, laboratory experiments
Conscious States	Mental Workload	Presence/Immersion
Unconscious States	Working Memory	Attentional orienting
Physical States	Continuous estimations from EEG data	Attentional-orienting ERPs
Body and World	Not considered	Conceptually and experimentally considered
States Manifestation	Task-induced	Stimulus-evoked
Simulator-Based Task	Synthetic and quasi-naturalistic simulated tasks and environment	Different degrees of quasi-naturalistic simulated tasks and environment
Primary "Within" Contributions	Theoretical, Methodological, Empirical	Theoretical, Methodological, Empirical, Practical
Primary "Between" Contributions	Methodological, Empirical	Theoretical, Methodological, Empirical

1.5.1 Chapter 2 - Essay #1 - Neurophysiological measurement of mental states for human-computer interaction: a scoping review

1.5.1.1 Introduction

Assessing cognitive and mental states plays a crucial role in neuroscience as it offers valuable insights into human brain function and behavior. However, the methods used for assessing these states often encounter limitations when applied to real-world tasks. Our current understanding of the brain and the hypotheses tested in controlled environments have limited generalizability to naturalistic settings (Matusz et al., 2019; Nastase et al., 2020; Stangl et al., 2023). Nevertheless, researchers in the field of neuro-adaptive technologies, specifically brain-computer interfaces (BCI), have developed methods for estimating mental states that have practical applications in complex data collection environments. BCI research combines mental state estimation within (quasi)naturalistic environments, enabling real-time monitoring and adaptation based on the user's brain signals (Zander & Kothe, 2011), which could inform our endeavor into mental state estimation in HCI and IS.

1.5.1.2 Objectives and Methodology

This review seeks to enhance the comprehension of neuro-adaptive system design and approaches in human-computer interaction (HCI) by conducting a scoping review. The review explores current research on neurophysiological measures and artificial intelligence techniques employed for mental state estimation in neuro-adaptive systems. The objectives of this paper are threefold: (i) to provide a comprehensive overview of the domain, tasks, methodologies, and psychophysiological inference utilized in neuro-adaptive system research, (ii) to develop a descriptive understanding of neuro-adaptive systems in naturalistic and quasi-naturalistic environments, and (iii) to identify gaps and challenges within the inherently multidisciplinary body of literature.

Given the emerging nature of the field and its fragmented presence across various domains, a scoping review methodology was adopted for this study. Following established guidelines for scoping reviews (Arksey & O'Malley, 2005; Daudt et al., 2013; Pham et al., 2014), this essay encompasses a broad spectrum of related disciplines, including brain-computer interfaces (BCI), neuroergonomics, medicine, and IS. The review specifically focuses on the utilization of neurophysiological measurements for assessing mental states.

A conceptual framework was built by combining a Design Science Research framework and a BCI design research framework to map out the domain, tasks, methodologies, and psychophysiological inference used in neuro-adaptive system research and to understand these systems in naturalistic environments. The record analysis will be done through a conceptual framework derived from Gregor and Hevner (2013), Hevner et al. (2004), Mason and Birch (2003), and Venable (2006). We treat the scientific contribution of the records as constructed artifacts and study them considering four three that influence the resulting technique: (i) the problem space, (ii) the knowledge space, and (iii) the solution space.

1.5.1.3 Contributions

The literature review identified a total of 36 empirical studies that met the inclusion criteria. Utilizing the three-dimensional framework that encompasses the problem, knowledge, and solution spaces, the findings of the review offer a comprehensive overview of the current state of research in the field. The analysis encompasses the initial problem that serves as the foundation for the development of a neuro-adaptive system, as well as the knowledge acquired throughout the research process, leading to the creation of the evaluated artifact or solution.

Our review offers three significant contributions. Firstly, it provides a comprehensive overview of the problems that drive the design of neuro-adaptive systems. Secondly, we present a thorough analysis of the functional components of neuro-adaptive systems and describe the relationships between different design choices and the corresponding psychophysiological inferences. This comprehensive examination enhances our understanding of how these systems operate and can guide future design decisions. Lastly, we consolidate the challenges encountered in neuro-adaptive systems research within the HCI field and propose guidelines and opportunities for future research. These include precise problem definitions, detailed descriptions of neurophysiological inferences, and the use of machine learning algorithms. By offering this guidance, we aim to assist IS researchers and designers. In summary, our review contributes by providing an extensive overview of the problems, conducting a comprehensive analysis of functional components, and offering guidelines for future research in developing neuro-adaptive systems.

1.5.2 Chapter 3 - Essay #2 - End-to-end deep learning approaches to mental workload classification using electroencephalography in HCI

1.5.2.1 Introduction

During demanding cognitive tasks like operating advanced aeronautical vehicles human information processing capabilities tend to decline due to mental workload limitations. This decline in processing abilities leads to reduced attention and narrowing the ability to process incoming sensory and information integration, subsequently impacting situational awareness and effective decision-making skills (Wickens, 2002).

This empirical paper adopts a task-induced perspective to investigate mental workload during a digital simulator task. The methodological approach involves manipulating continuous electrical patterns in the brain using a validated task and aims to generalize these patterns to a novel task, specifically a flight training task.

Machine learning is a common technique for inferring mental states from continuous electrical patterns (Craik et al., 2019; Muller et al., 2008). We surveyed the literature on state-of-the-art techniques for mental workload decoding using continuous signals. The result showed an important gap in end-to-end techniques. End-to-end decoding approaches attempt to classify brain signal with no manual feature engineering. Other authors have signaled the presence of this gap for EEG analysis in general (Roy, Banville, et al., 2019).

An end-to-end process in the context of deep learning for mental workload estimation describes a process that takes raw EEG signal data, processes these data, derives discriminant and invariant features, then provides a classification of the target state as a complete functional solution. During our review of the literature regarding deep learning approaches to mental workload estimation, we discovered that most research approaches fail to leverage a major strength of the method: the ability to learn discriminative features and produce classifications directly from the raw EEG signal. Instead, most deep learning approaches toward estimating mental workload from EEG signals transform these data into features within the time or frequency domain.

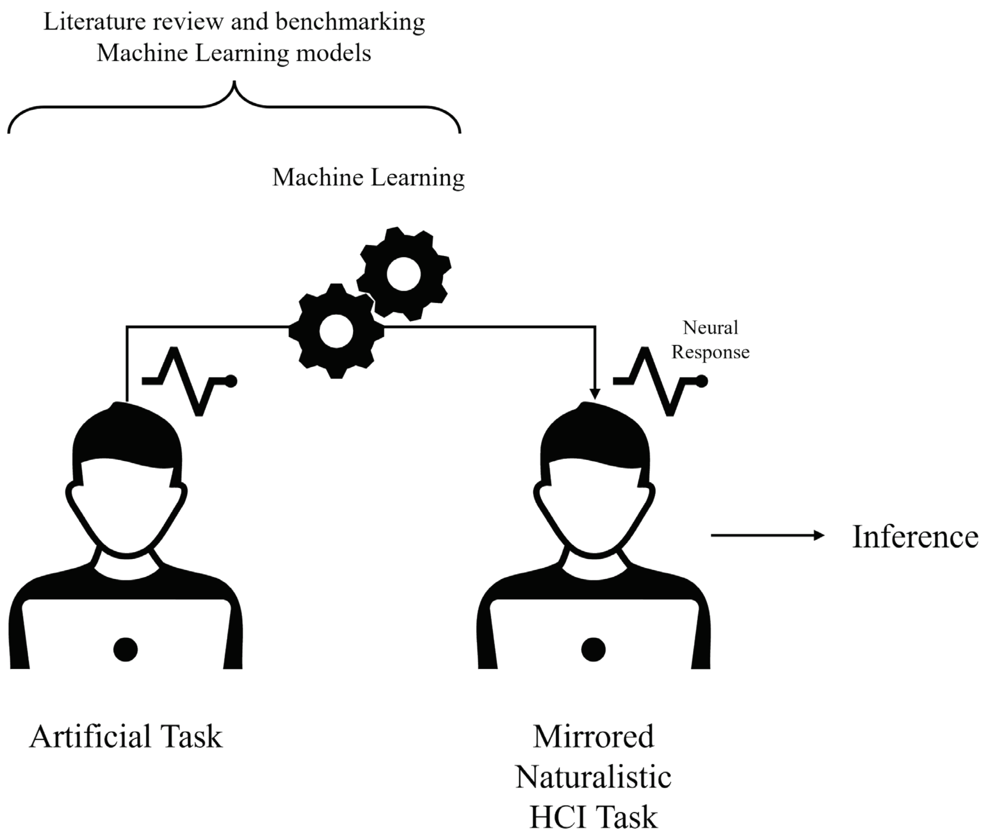
1.5.2.2 Objectives and Methodology

Therefore, our overarching research objective in this essay is to create a deep learning model based on end-to-end methods for estimating mental workload during flight training within a high-fidelity simulator. The manuscript seeks to answer the question, "to what extent is it possible to estimate mental workload during naturalistic HCI tasks based on neurophysiological signal data using an end-to-end deep learning process?". In order to answer the research question, the objectives of this work are threefold: (1) Benchmark end-to-end deep-learning models for mental workload classification; (2) Develop a mental workload classifier that achieves high classification performance; (3)

Estimate mental workload during a simulator task and replicate past empirical findings related to the relationship between task complexity, mental workload, and performance. Figure 7 represents a high-level depiction of the research approach.

Figure 7

Schematic methodological strategy for essay #2 – chapter 3 using a task-induced paradigm



Note. Within-subject mental state decoding with machine learning, the same participant performs the artificial and the naturalistic HCI tasks.

Eleven pilots participants performed within-subject experimental design composed of two experimental tasks. First, a synthetic and validated task consisting of an n-back meant to manipulate mental workload. The task has been used in a high number of studies to induce different levels of mental workload and is widely accepted for mental workload estimation (Baldwin & Penaranda, 2012; Hefron et al., 2018; Kim et al., 2014; Kuanar et al., 2018; Saadati et al., 2020). An ecologically valid flight task was designed to induce different levels of mental workload mirroring n-back difficulty levels through the manipulation of maneuver difficulty.

Informed by the ML and neuroscience literature, we made model design choices specific to EEG signal data and systematically assessed those choices in terms of performance and neurophysiological plausibility. We followed good practice in using machine learning for neuroimaging (Kohoutová et al., 2020). Design choices were systematically justified and tested. Moreover, models and features were assessed for their neurophysiological plausibility. We benchmarked several deep learning models and selected the two best-performing architectures (i.e., FCN, a fully convolutional neural network and, ResNet, a residual network).

Informed by the machine learning (ML) and neuroscience literature, design choices specific to EEG signal data were taken, and we present a systematic assessment of those choices. The two selected classifier models achieved an average accuracy of .933 (± .054) and .917 (± .074) for FCN and ResNet, respectively. We validate our models through rigorous assessments of their neurophysiological plausibility, robustness, and reliability. We then deployed these two classifier models on previously unobserved EEG data, intending to duplicate empirical outcomes pertaining to the interplay between complexity, mental workload, and performance. Our classification findings revealed a potential inverse U-shaped relationship between complexity and mental workload.

1.5.2.3 Contributions

The primary contributions of this study encompass empirical, methodological, and theoretical aspects. From an empirical perspective, the ResNet and FCN end-to-end deep learning models demonstrate superior performance compared to existing baselines in estimating mental workload, highlighting the practicality and effectiveness of these models. Furthermore, the utilization of end-to-end deep learning models enables the application of transfer learning techniques, which enhance generalizability, reduce training time, and facilitate test-retest capability. Methodologically, a framework is proposed for benchmarking end-to-end deep learning models and conducting neurophysiological validation, thereby facilitating mental state estimation and future research in IS. Lastly, the study has theoretical implications for design science research and evaluation, emphasizing the importance of measuring constructs during tasks.

1.5.3 Chapter 4 - Essay #3 - Oriented-attention measurement in multisensory human-computer interaction using electroencephalography

1.5.3.1 Introduction

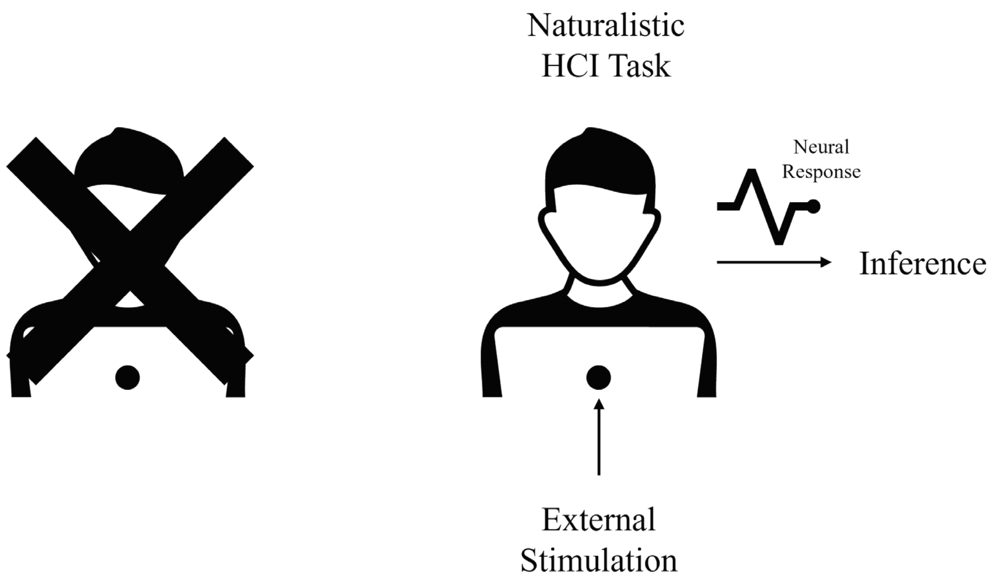
The careful consideration of the mechanisms of attention and multisensory integration is crucial for understanding the impact of HCI on users. In naturalistic tasks and environments, technology users are subjected to the salience of ongoing concurrent events while exerting executive control to maintain attentional resources on the task at hand (Matusz et al., 2019). However, the cognitive mechanisms involved in multisensory integration and its interplay with attention in real-world human-computer interaction remains little explored. Gaining insights into these processes and developing methods to study them in HCI tasks could significantly enhance our understanding of technology's implications on users at the cognitive mechanism level for IS research. Furthermore, it could offer innovative techniques for evaluating artifacts in design science research.

1.5.3.2 Objectives and Methodology

To elucidate the role of naturalistic environments, multisensory integration, and attention during HCI tasks, we build on a conceptual framework that bridges multisensory integration mechanisms with attention (Talsma et al., 2010). This framework demonstrates how multisensory integration is sensitive to both bottom-up and top-down attentional forces. Consequently, we develop a multisensory perturbation technique to trigger a drift of the allocation of attention from the HCI task toward a distractor to measure the orientation of attention. Figure 8 presents a high-level representation of the approach described. It this approach, we do not need a synthetic task to learn brain responses from, but we theoretically predict a neurophysiological response (i.e., evoked-related potentials) which is sensible to bottom-up and top-down factors. Finally, by utilizing multisensory microworlds, we iteratively increase naturalness to ensure that our measurement generalizes to quasi-naturalistic environments.

Figure 8

Schematic methodological strategy for essay #3 – chapter 4 using a stimulus-evoked paradigm



Note. This approach does not use a synthetic task before the naturalistic HCI task.

This essay aims to (1) better understand the role of intention orientation in multisensory simulation training at the neurophysiological level and (2) develop a novel measurement approach to the orientation of attention in a multisensory digital simulation context.

We designed two studies. The first study is between subject experimental design composed of 16 participants. The design was composed of 1-factor manipulating the presence of the synchronous movement of the haptic chair. We created a multisensory environment (i.e., auditory, visual, haptic) with a racing video game and a haptic chair mimicking the physics of the car. Auditory and visual modalities were kept constant. In addition, an intermittent and random pure-tone auditory stimulus was triggered to generate an ERP. The objective was to measure the allocation toward the distractor at the physical level by analyzing the brain response to this auditory stimulation. This study aims to reproduce past measurements to assess our capability to measure the orientation of attention in a multisensorial context using a traditional method, an unrelated auditory distractor.

However, this approach showed limitations in multisensory contexts. Multisensory events tend to strongly capture attentional resources and have a gating effect on unpredictable stimulation in non-dominant sensory modalities to the task. In study two, we design a salient distractor that fits characteristics for automatic integration in a multisensory context to address this limitation. Using this novel stimulation, we aim to assess the orientation of attention and performance in a racing training simulator. With a two-by-two within-subject experimental design, we are manipulating movement and attentional demands for 23 participants inside a quasi-realistic HCI simulation context in the laboratory.

1.5.3.3 Contributions

Drawing on a conceptual framework that bridges attention and multisensory integration (Talsma et al., 2010), this manuscript presents a methodological approach that enables the exploration of covert mechanisms during complex and dynamic tasks in quasi-naturalistic environments using a perturbation technique. The multisensory environment enhances attention orientation towards the task and reduces resource allocation for processing the perturbation. The results highlight the presence of a gating mechanism within a high multisensory HCI environment, where the brain prioritizes processing relevant sensory feedback crucial for task performance while inhibiting irrelevant distractions, even within the same sensory modalities. This finding aligns with previous research indicating that task demands, and cognitive processing influence late components of multisensory perturbation. In addition, the event-related potential (ERP) components are influenced by top-down attention. Therefore, the studies show that subjects allocate fewer resources to distractors when exposed to high multisensory environments and task sections requiring executive attention.

These studies make a valuable contribution to the field of attention in cognitive neuroscience by investigating the interplay between attention and multisensory contexts in quasi-naturalistic situations. The findings suggest that the orientation of attention to task-relevant but goal-irrelevant perturbation might show a reduction of attention drift when sensory modalities are added. This highlights the capacity of multisensory environments to effectively capture attentional resources. It offers conceptual advancement by better understanding how attention and multisensory integration interact in quasi-naturalistic environments.

Furthermore, the approaches provide insights into the underlying cognitive mechanisms associated with constructs such as immersion, absorption, or presence, offering implications for IS and HCI research. The research methodology used allows for an in-depth examination of relevant psychological constructs at the cognitive level, allowing IS researchers to gain a deeper understanding of their implications.

Additionally, this study contributes to IS research by demonstrating how cognitive neuroscience can be used to inform and evaluate artifact designs. By addressing the evaluation methods through the lens of attentional orientation, the study aligns with the core principles of NeuroIS. The practical significance of this research lies in the identification of system design choices that impact attention in multisensory contexts, positively influencing the allocation of attentional resources. The studies also highlight the impact of multisensory simulation and microworld design on information processing, providing valuable insights for artifact designers constructing digital simulations for training purposes through the integration of multiple sensory modalities.

References

Alzahab, N. A., Apollonio, L., Di Iorio, A., Alshalak, M., Iarlori, S., Ferracuti, F., Monteriù, A., & Porcaro, C. (2021). Hybrid deep learning (hDL)-based brain-computer interface (BCI) systems: a systematic review. *Brain sciences*, *11*(1), 75.

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, *8*(1), 19-32.

Baker, M. (2016). Reproducibility crisis. *Nature*, *533*(26), 353-366.

Baldwin, C. L. (2019). Neuroergonomics of Simulators and Behavioral Research Methods. In *Neuroergonomics* (pp. 49-53). Elsevier.

Baldwin, C. L., & Penaranda, B. N. (2012). Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. *Neuroimage*, *59*(1), 48-56. https://doi.org/10.1016/j.neuroimage.2011.07.047

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., Poeppel, D., & Dikker, S. (2019). Brain-to-Brain Synchrony and Learning Outcomes Vary by Student-Teacher Dynamics: Evidence from a Real-world Classroom Electroencephalography Study. *J Cogn Neurosci*, *31*(3), 401-411. https://doi.org/10.1162/jocn_a_01274

Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, *327*(8476), 307-310.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, *9*(2-3), 171-184.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (2007). *Handbook of psychophysiology*. Cambridge university press.

Chudnoff, E. (2015). *Cognitive phenomenology*. Routledge.

Churchland, P. S. (1989). *Neurophilosophy: Toward a unified science of the mind-brain*. MIT press.

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251)

Conrad, C., & Bailey, L. (2020). What Can NeuroIS Learn from the Replication Crisis in Psychological Science? In *Information Systems and Neuroscience* (pp. 129-135). Springer.

Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering*, *16*(3), Article 031001. https://doi.org/10.1088/1741-2552/ab0ab5

Daudt, H. M., van Mossel, C., & Scott, S. J. (2013). Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC medical research methodology*, *13*(1), 1-9.

De Guinea, A. O., & Markus, M. L. (2009). Why break the habit of a lifetime? Rethinking the roles of intention, habit, and emotion in continuing information technology use. *MIS quarterly*, 433-444.

De Guinea, A. O., Titah, R., & Léger, P.-M. (2013). Measure for measure: A two study multi-trait multi-method investigation of construct validity in IS research. *Computers in Human Behavior*, *29*(3), 833-844.

De Winter, J., van Leeuwen, P. M., & Happee, R. (2012). Advantages and disadvantages of driving simulators: A discussion. Proceedings of measuring behavior,

Demazure, T., Karran, A., Léger, P.-M., Labonté-LeMoyne, É., Sénécal, S., Fredette, M., & Babin, G. (2021). Enhancing Sustained Attention. *Business & Information Systems Engineering*, *63*(6), 653-668. https://doi.org/10.1007/s12599-021-00701-3

Dimoka, A., Davis, F. D., Gupta, A., Pavlou, P. A., Banker, R. D., Dennis, A. R., Ischebeck, A., Müller-Putz, G., Benbasat, I., & Gefen, D. (2012). On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS quarterly*, 679-702.

Dimoka, A., Pavlou, P. A., & Davis, F. D. (2011). Research commentary—NeuroIS: The potential of cognitive neuroscience for information systems research. *Information Systems Research*, *22*(4), 687-702.

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature neuroscience*, *8*(12), 1643-1646.

Frey, J., Hachet, M., & Lotte, F. (2018). Recent Advances in EEG-Based Neuroergonomics for Human–Computer Interaction. *Neuroergonomics*, 275.

Gómez-Zará, D., Schiffer, P., & Wang, D. (2023). The promise and pitfalls of the metaverse for science. *Nature Human Behaviour*, 1-4.

Gonsalvez, C. J., & Polich, J. (2002). P300 amplitude is determined by target-to-target interval. *Psychophysiology*, *39*(3), 388-396.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, *37*(2).

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in cognitive sciences*, *14*(1), 40-48.

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature reviews neuroscience*, *7*(7), 523-534.

Hefron, R., Borghetti, B., Schubert Kabban, C., Christensen, J., & Estepp, J. (2018). Cross-Participant EEG-Based Assessment of Cognitive Workload Using Multi-Path Convolutional Recurrent Neural Networks. *Sensors*, *18*(5), 1339.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, *28*(1), 75-105.

Johnson, A., & Proctor, R. (2013). *Neuroergonomics: A cognitive neuroscience approach to human factors and ergonomics*. Springer.

Karran, A. J., Demazure, T., Léger, P.-M., Labonte-LeMoyne, E., Sénécal, S., Fredette, M., & Babin, G. (2018). Towards a hybrid passive BCI for the modulation of sustained attention using EEG and fNIRS [Abstract]. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/conf.fnhum.2018.227.00115

Kim, J., Kim, M.-K., Wallraven, C., Kim, S.-P., & Ieee. (2014). *Across-subject estimation of 3-back task performance using EEG signals*.

Kirwan, C. B., Vance, A., Jenkins, J. L., & Anderson, B. B. (2023). Embracing brain and behaviour: Designing programs of complementary neurophysiological and behavioural studies. *Information Systems Journal*.

Kohoutová, L., Heo, J., Cha, S., Lee, S., Moon, T., Wager, T. D., & Woo, C.-W. (2020). Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nature protocols*, *15*(4), 1399-1435.

Kotchoubey, B., Tretter, F., Braun, H. A., Buchheim, T., Draguhn, A., Fuchs, T., Hasler, F., Hastedt, H., Hinterberger, T., Northoff, G., Rentschler, I., Schleim, S., Sellmaier, S., Tebartz Van Elst, L., & Tschacher, W. (2016). Methodological Problems on the Way to Integrative Human Neuroscience [Review]. *Frontiers in Integrative Neuroscience*, *10*. https://doi.org/10.3389/fnint.2016.00041

Kuanar, S., Athitsos, V., Pradhan, N., Mishra, A., & Rao, K. (2018). Cognitive Analysis of Working Memory Load from EEG, by a Deep Recurrent Neural Network. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2016). Understanding Minds in Real-World Environments: Toward a Mobile Cognition Approach. *Front Hum Neurosci*, *10*, 694. https://doi.org/10.3389/fnhum.2016.00694

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2019). Mobile EEG identifies the re-allocation of attention during real-world activity. *Sci Rep*, *9*(1), 15851. https://doi.org/10.1038/s41598-019-51996-y

Léger, P.-M. (2006). Using a simulation game approach to teach enterprise resource planning concepts. *Journal of Information Systems Education*, *17*(4), 441.

Léger, P.-M., Charland, P., Feldstein, H. D., Robert, J., Babin, G., & Lyle, D. (2011). Business simulation training in information technology education: Guidelines for new approaches in IT training. *Journal of Information Technology Education*, *10*(1), 39-53.

Léger, P.-M., Sénecal, S., Courtemanche, F., de Guinea, A. O., Titah, R., Fredette, M., & Labonte-LeMoyne, É. (2014). Precision is in the eye of the beholder: Application of eye fixation-related potentials to information systems research.

Léger, P., Robert, J., Babin, G., Pellerin, R., & Wagner, B. (2007). ERPsim. *ERPsim Lab (erpsim. hec. ca), HEC Montreal, QC*.

Loos, P., Riedl, R., Müller-Putz, G. R., Vom Brocke, J., Davis, F. D., Banker, R. D., & Léger, P.-M. (2010). NeuroIS: neuroscientific approaches in the investigation and development of information systems. *Business & Information Systems Engineering*, *2*(6), 395-401.

Lotte, F., & Roy, R. N. (2019). Brain–Computer Interface Contributions to Neuroergonomics. In *Neuroergonomics* (pp. 43-48). Elsevier.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. MIT Press.

Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The curious incident of attention in multisensory integration: bottom-up vs. top-down. *Multisensory Research*, *29*(6-7), 557-583.

Mason, S. G., & Birch, G. E. (2003). A general framework for brain-computer interface design. *IEEE Trans Neural Syst Rehabil Eng*, *11*(1), 70-85. https://doi.org/10.1109/TNSRE.2003.810426

Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019). Are we ready for real-world neuroscience? , *31*(3), 327-338.

Mehta, R., & Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work [Review]. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00889

Midha, S., Maior, H. A., Wilson, M. L., & Sharples, S. (2021). Measuring Mental Workload Variations in Office Work Tasks using fNIRS. *International Journal of Human-Computer Studies*, *147*, 102580. https://doi.org/10.1016/j.ijhcs.2020.102580

Mingers, J. (2001). Combining IS research methods: towards a pluralist methodology. *Information Systems Research*, *12*(3), 240-259.

Müller-Putz, G. R., Riedl, R., & Wriessnegger, S. C. (2015). Electroencephalography (EEG) as a Research Tool in the Information Systems Discipline: Foundations, Measurement, and Applications. *CAIS*, *37*, 46.

Muller, K. R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. *J Neurosci Methods*, *167*(1), 82-90. https://doi.org/10.1016/j.jneumeth.2007.09.022

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *Neuroimage*, *222*, 117254.

Northoff, G. (2018). *The spontaneous brain: from the mind-body to the world-brain problem*.

Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Research synthesis methods*, *5*(4), 371-385.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature reviews neuroscience*, *18*(2), 115-126.

Polich, J. (1989). Habituation of P300 from auditory stimuli. *Psychobiology*, *17*(1), 19-28.

Polich, J. (1998). P300 clinical utility and control of variability. *Journal of Clinical Neurophysiology*, *15*(1), 14-33.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol*, *118*(10), 2128-2148. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J., & Heine, M. R. (1996). P300 topography and modality effects from a single-stimulus paradigm. *Psychophysiology*, *33*(6), 747-752.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological psychology*, *41*(2), 103-146.

Pope, A. T. B., Edward H.; Bartolonne, Debbie S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task.

Riedl, R., Davis, F. D., & Hevner, A. R. (2014). Towards a NeuroIS research methodology: intensifying the discussion on methods, tools, and measurement. *Journal of the Association for Information Systems*, *15*(10), I.

Riedl, R., Fischer, T., Léger, P.-M., & Davis, F. D. (2020). A decade of NeuroIS research: progress, challenges, and future directions. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *51*(3), 13-54.

Riedl, R., & Léger, P.-M. (2016). Fundamentals of NeuroIS. *Studies in Neuroscience, Psychology and Behavioral Economics. Springer, Berlin, Heidelberg*.

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *49*(3), 329-359.

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, *16*(5), Article 051001. https://doi.org/10.1088/1741-2552/ab260c

Roy, Y., Hubert, B., Isabela, A., Alexandre, G., & Jocelyn, F. (2019). Deep learning-based electroencephalography analysis: a systematic review. *arXiv preprint arXiv:1901.05498*.

Saadati, M., Nelson, J., & Ayaz, H. (2020). Multimodal fNIRS-EEG Classification Using Deep Learning Algorithms for Brain-Computer Interfaces Purposes. In H. Ayaz (Ed.), *Advances in Neuroergonomics and Cognitive Engineering* (Vol. 953, pp. 209-220). https://doi.org/10.1007/978-3-030-20473-0_21

Schwarz, N., & Clore, G. L. (2007). Feelings and phenomenal experiences.

Spiers, H. J., & Maguire, E. A. (2007). Decoding human brain activity during real-world experiences. *Trends Cogn Sci*, *11*(8), 356-365. https://doi.org/10.1016/j.tics.2007.06.002

Stangl, M., Maoz, S. L., & Suthana, N. (2023). Mobile cognition: imaging the human brain in the 'real world'. *Nature reviews neuroscience*, 1-16.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences*, *14*(9), 400-410.

Tams, S., Hill, K., de Guinea, A. O., Thatcher, J., & Grover, V. (2014). NeuroIS-alternative or complement to existing methods? Illustrating the holistic effects of neuroscience and self-reported data in the context of technostress research. *Journal of the Association for Information Systems*, *15*(10), 723.

Venable, J. (2006). The role of theory and theorising in design science research. Proceedings of the 1st International Conference on Design Science in Information Systems and Technology (DESRIST 2006),

vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a neurois research agenda with four areas of societal contributions. *European Journal of Information Systems*, *29*(1), 9-24.

vom Brocke, J., & Liang, T.-P. (2014). Guidelines for Neuroscience Studies in Information Systems Research. *Journal of management information systems*, *30*(4), 211-234. https://doi.org/10.2753/MIS0742-1222300408

Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L. A., & Coughlin, J. F. (2010). The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics*, *53*(3), 404-420.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science*, *11*(4), 128-133.

Zander, T. O., & Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of neural engineering*, *8*(2), 025005..

Chapter 2
Essay #1 - Neurophysiological Measurement of Mental States for Human-Computer Interaction and Adaptation: A Scoping Review

Abstract

Background: Naturalistic HCI tasks and environments impose limitations on the use of neurophysiological methods. Cognitive and mental state assessment is essential in neuroscience, but traditional methods have limitations when applied to real-world tasks. Researchers in neuro-adaptive technologies and brain-computer interfaces have developed methods for mental state estimation in complex data collection environments.

Objective: This review aims to increase the understanding of mental state estimation by conducting a scoping review that investigates the current research on neurophysiological measures and artificial intelligence techniques for mental state estimation in adaptive environments to inform NeuroIS research.

Method: Due to the emergence of the field and its scatteredness across domains, we employed a scoping review methodology. A conceptual framework was built by combining a Design Science Research framework and a BCI design research framework to map out the domain, tasks, methodologies, and psychophysiological inference used in neuro-adaptive system research and to understand these systems in naturalistic environments.

Results: The literature search resulted in 36 studies that met the inclusion criteria. By employing a three-dimensional framework (i.e., problem, knowledge, solution) to characterize the literature, the results provide a comprehensive overview of the current state of research, ranging from the initial problem that justifies the need for a neuro-adaptive system to the final evaluated artifact.

Conclusions: The scoping review highlights challenges encountered by neuro-adaptive systems in HCI, including problem definitions, descriptions of neurophysiological

inferences, and the utilization of machine learning algorithms. We provide guidance for researchers and designers within this domain and further work to enable mental state inference in naturalistic IS tasks.

2.1 Introduction

Cognitive and mental state assessment is crucial in neuroscience as it provides vital insights into human brain function and behavior. However, the methods used to assess cognitive and mental states face limitations when applied to real-world tasks. Our understanding of the brain and the hypotheses developed and tested in controlled environments generalize little to naturalistic ones (Matusz et al., 2019; Nastase et al., 2020; Stangl et al., 2023). However, researchers in the field of neuro-adaptive technologies, or brain-computer interfaces (BCI), have developed methods for mental state estimation that have real-world applications in complex data collection environments. BCI research combines mental state estimation with naturalistic human-computer interaction (HCI) to monitor and adapt in real-time to the user's brain signals (Zander & Kothe, 2011). vom Brocke et al. (2020) anticipate that neuro-adaptive technologies might contribute to the development of innovative artifacts and methods in NeuroIS.

Naturalistic HCI tasks and environments impose limitations on the use of neurophysiological measures to estimate mental state, but the literature on BCI in HCI can shed light on this challenge. Those systems and mental state estimation techniques rely on the neurophysiological signal. In the case of electroencephalography (EEG) in naturalistic tasks, unconstrained users' mental processes, body movements, and environments cause other noises in the electrical signal recorded, which requires developing novel techniques. For example, Kline et al. (2015) created a technique to isolate gait artifacts induced in EEG signals while walking using a nonconductive layer on top of the sensors in addition to artifact removal techniques. Algorithmic techniques to improve the signal-to-noise ratio in the signal are also employed. For example, Independent Component Analysis (ICA) is the most popular method to remove artifacts (Gorjan et al., 2022). Moreover, feature extraction and machine learning techniques can also be leveraged. For example, Appriou et al. (2018) recognized the need for more robust classification techniques for workload estimation in HCI. To do so, the authors benchmarked machine learning techniques from the BCI domains that achieved important performance.

A BCI is a cyclic process where neural signals are acquired, processed, and translated into output commands that control an external device or application; this enables users to interact with their environment using neurophysiological activity alone (Van Gerven et al., 2009). The field emphasizes robust inference methods and evaluates technology through user interaction. BCI research is often referred to as live biofeedback (Lux et al., 2018), neuro-adaptive systems (vom Brocke et al., 2020), BCIs (Mason & Birch, 2003), brain-machine interfaces (Nicolelis & Lebedev, 2009), and physiological computing (Fairclough, 2009). For the purpose of this review, we shall adopt the term "neuro-adaptive system(s)" to encompass the entirety of research within the BCI field.

Neuro-adaptive systems are moving from experimental laboratory prototypes into real-world applications. Advances in real-time processing of neurophysiological data and Artificial Intelligence (AI) applications for signal analysis (Roy et al., 2019) coupled with reduced sensor costs and data-dense environments (e.g., industry 4.0, healthcare) are facilitating this movement, and opportunities have started to emerge in the industry (Whelan et al., 2018), healthcare (Gu et al., 2021), and aeronautics (Lotte & Roy, 2019). Neuro-adaptive systems have been proposed to enhance human and system interaction in HCI (Zander & Kothe, 2011; Zander et al., 2010). Information systems research also shows interest in recent calls for research (vom Brocke et al., 2020), literature review (Lux et al., 2018), and design science research on neuro-adaptive artifacts (Demazure et al., 2021; Toreini et al., 2022). Those fields explored diverse mental states, such as visual attention allocation (Toreini et al., 2022), vigilance (Di Flumeri et al., 2019), mental workload (Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016), or affective state (Govindarajan et al., 2018). One of the main challenges of applied neuro-adaptive systems in HCI is that this literature is scattered and emerging across different research domains (e.g., neuroscience, ergonomics, engineering, and information systems).

We aim to increase the current understanding of the design of neuro-adaptive systems (i.e., users, tasks, and environments) and design approaches (e.g., algorithms and psychophysiological inferences) in HCI, which could also inform mental state estimation in naturalistic IS tasks. To address this goal, we conducted a scoping review that answers the following research question: “What is the current state of research on the use of neurophysiological measures and artificial intelligence techniques for mental state estimation in HCI within adaptive environments?” For this review, we focused on artifacts with a certain degree of practical purpose in HCI. Our objectives are (i) to map out the domain, tasks, methodologies, and psychophysiological inference utilized in neuro-adaptive system research, (ii) to build a descriptive understanding of neuro-adaptive systems in naturalistic and quasi-naturalistic environments, and (iii) to extract gaps and challenges from an inherently multidisciplinary body of literature. To achieve this, we built a conceptual framework by blending a Design Science Research framework (Gregor & Hevner, 2013; Hevner et al., 2004; Venable, 2006) and BCI design research framework (Mason & Birch, 2003; Van Gerven et al., 2009).

We address the call for research presented by Riedl and Léger (2016) calling for the exploration of users' neurophysiological data in adaptive systems, as they argue that such systems can yield favorable outcomes in terms of health and performance. This avenue of investigation presents a distinctive opportunity for IS artifacts to incorporate users' emotions and cognition, thereby enhancing the capabilities of both the artifact and the user. Moreover, vom Brocke et al. (2020) anticipate that neuro-adaptive systems represent a promising field for the future of IS, with the potential for significant societal contributions. Consequently, an in-depth understanding of the existing literature on applied neuro-adaptive systems is crucial for advancing future design knowledge and facilitating practical applications.

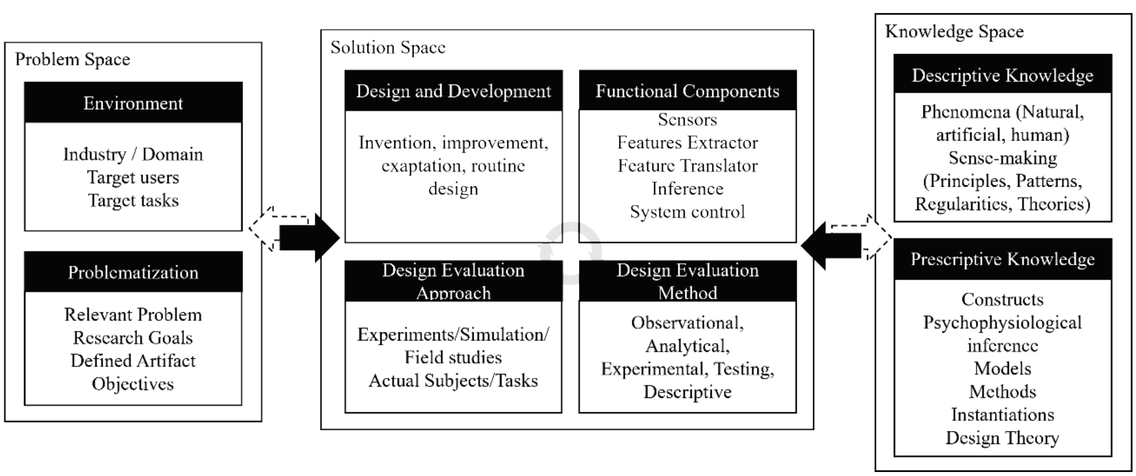
Our review makes three main contributions. First, it provides a broad overview of the design of neuro-adaptive systems. Second, we present a comprehensive analysis of the functional components of neuro-adaptive systems and the relationships between different design choices and psychophysiological inferences. Third, we consolidate the challenges found in the current literature and suggest guidelines and opportunities for future research. The remainder of the manuscript is organized as follows: We introduce our analytical framework and outline the methodology used. Next, we present the descriptive results of the review before synthesizing the findings and identifying the challenges in the discussion. Finally, we address the limitations and offer conclusions.

2.2 Analytical Framework

To answer our research question and specify the scope of this review, we build upon the framework from Design Science Research (DSR) (Gregor & Hevner, 2013; Hevner et al., 2004; Venable, 2006). The DSR framework aims to inform the creation, evaluation, and dissemination of innovative artifact-building on design science and behavioral research. Our framework is built around three spaces: the problem space, the knowledge space, and the solution space (Figure 9).

Figure 9

Conceptual framework for the data charting, collating, and summarizing of neuro-adaptive artifacts



The *problem space* (Table 2) is defined by the researcher's understanding of the problem (Venable, 2006). In the current case, it comprises the environment and the nature of the interaction within the socio-technical system. Within this space, users and their characteristics, tasks, organizations, or the technological ecosystem are all parts of the structure that will motivate the need for a neuro-adaptive artifact and the problem for which the researcher needs to provide a solution (Hevner et al., 2004). In essence, the problem space represents the situated knowledge justifying the need for the neuro-adaptive artifact and directly influences the solution space.

The *problem space* has two categories, the *environment* and *problematization*. The first category represents the environment from which the problem and the need for an artifact emerges. System designers build neuro-adaptive systems for prospective users, tasks, and operating environments. For instance, neuro-adaptive systems are researched in aeronautics, notably to support operators during air traffic control tasks (Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016). Therefore, effective designs should consider contextual users' inherent characteristics, the task's nature, and the environment's constraints (Mason & Birch, 2003). These elements become critical when system designers plan to deploy the artifact in naturalistic operating environments. The second category, problematization, represents the motivation of the researchers concerning the created and evaluated artifact. As such, the researchers should define the problem and communicate the artifact's research objective and goals (Gregor & Hevner, 2013). Problematization is essential to motivate the artifact's need and define its purpose, scope, and relevance. In the case of neuro-adaptive artifacts, it can be the factors from which the problem emerges and the outcomes it leads to justifying the need to construct an artifact. For instance, surveillance tasks on the computer pressure the attentional resources and favor attentional decrement, which can lead to critical mistakes (Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016; Demazure et al., 2019; Di Flumeri et al., 2019).

Table 2

Problem space categories, components, and definitions

Category	Components	Definition	Source
Environment	Target Users	The expected user for whom the neuro-adaptative system is designed.	Mason & Birch, 2003
	Target Task	The expected task (and their characteristics) the neuro-adaptive system is designed to support.	Mason & Birch, 2003
	Organization / Domain	Organization/Domain from which the problem emerges	Hevner et al., 2004
Problematization	Problem Statement	The relevant problem the artifact aims to resolve	Gregor & Hevner, 2013
	Research Objective / Research Question	The objective of the research	
	Artifact Goals	The objective of the created and evaluated artifact	

The *knowledge space* (Table 3) represents the building blocks necessary for constructing the artifact (Hevner et al., 2004). Before creating neuro-adaptive systems, researchers and designers inquire "what do we know already? From what existing knowledge can we draw?" (Gregor & Hevner, 2013, p. 343). It comprises *descriptive* and *prescriptive knowledge* (Gregor & Hevner, 2013). Descriptive knowledge is the theories and patterns derived from the literature to inform the solution space. For example, psychophysiological phenomena are often complex given that psychological measures do not respect one-to-one relationships with mental states (Fairclough, 2009), forcing neuro-adaptive systems researchers to ground artifacts with neurophysiological knowledge to ensure plausibility. However, those inferences are often correlated with mental state theories, forcing researchers to anchor solutions to descriptions of the cognitive construct or phenomena of interest. Prescriptive knowledge embodies the "know-how", which relates to past artifacts, algorithms, or methods used to solve similar research problems in the literature. As defined by (Gregor & Hevner, 2013), "prescriptive knowledge concerns artifacts designed by humans to improve the natural world."

Table 3

Knowledge space categories, components, and definitions

Category	Components	Definition	Source
Descriptive Knowledge	Phenomena	"Human-related phenomena are composed of observations, classifications, measurements, and the cataloging of these descriptions into accessible form"	Gregor & Hevner, 2013
	Sense-Making	"Sense-making is represented by natural laws, principles, regularities, patterns, and theories"	
Prescriptive Knowledge	Constructs	"Constructs, which provide the vocabulary and symbols used to define and understand problems and solutions; for example, the constructs of “entities” and “relationships” in the field of information modeling"	Gregor & Hevner, 2013
	Models (and frameworks)	"Models are designed representations of the problem and possible solutions. [...] Models corresponds to […] the abstract blueprint of an artifact’s architecture, which show an artifact’s components and how they interact."	
	Methods	"Methods are algorithms, practices, and recipes for performing a task"	
	Instantiations	“The physical realizations that act on the natural world, such as an information system that stores, retrieves, and analyzes customer relationship data. Instantiations can embody design knowledge, possibly in the absence of more explicit description. The structural form and functions embodied in an artifact can be inferred to some degree by observing the artifact."	
	Design Theory	"Which is an abstract, coherent body of prescriptive knowledge that describes the principles of form and function, methods, and justificatory theory that are used to develop an artifact or accomplish some end"	
	Psychophysiological inference	Describe the neurophysiological mechanisms' relationship with a psychological construct	Mason & Birch, 2003

The *solution space* (Table 4) is the process of building and evaluating the artifact. The literature targeted for this review often presents the produced neuro-adaptive artifact as the primary research contribution, which can take many forms (e.g., algorithm, software, physical machine). The solution space is often an iterative process between the design and development of the artifact. Its evaluation can be viewed as the treatment for the problem identified.

Table 4

Solution space (design and development, design evaluation approach, and design evaluation method) categories, components, and definitions

Category	Components	Definition	Source
Design and Development	Invention	"Invent new solutions for new problems"	Gregor & Hevner, 2013
	Improvement	"Develop new solutions for known problems"	
	Exaptation	"Extend known solution new problems"	
	Routine Design	"Apply known solutions to known problems"	
Design Evaluation Method	Actual Subjects	The actual user the neuro-adaptative the system is evaluated with.	Mason & Birch, 2003
	Actual Task	The actual task the neuro-adaptive system supports during evaluation.	
	Actual Operating Environment	refers to the physical environment, objects, and people around which the user interacts with the BCI during evaluation.	
Design Evaluation Method	Observational	Case Study, Field Study	Hevner et al., 2004
	Analytical	Static Analysis, Architecture Analysis, Optimization, Dynamic Analysis	
	Experimental	Controlled Experiment, Simulation	
	Testing	(Black Box) Testing, Structural (White Box) Testing	
	Descriptive	Informed Argument, Scenarios	

Within the solution space, *functional components* (Table 5) refer to the building blocks of the neuro-adaptive artifact. It represents the characteristics of the interaction human-machine, the type of interaction that might directly influence the sensor selection, the neurophysiological mechanisms targeted, the features extracted from the signal, the translation of these features, the dynamic of control between the operator and the machine, and how the machine adapts to its user (Mason & Birch, 2003). They represent the minimal components that characterize neuro-adaptive systems.

Table 5

Functional components and definitions

Category	Components	Definition	Source
Functional components	Sensors	Supplier/Type of the electrodes used during the evaluation.	Mason & Birch, 2003
	Feature extractor	Description of the method used to create the features vector, or indexes/scores used as inputs	
	Feature translator	Description of the method used to translate the feature vector into logical control signals	
	Neurophysiological mechanisms	Description of the neurophysiological mechanisms or processes the user uses to control the neuro-adaptive system (pattern, physiological response, etc.)	
	Neuro-adaptive systems control	Description of the control mechanisms (e.g., feedback, retroaction, adaptation, allocation)	

In summary, we employ this conceptual framework to map and characterize the literature on neuro-adaptive system design within HCI. The framework comprises the solution space, which arises from the interplay between the problem and knowledge spaces. For this scoping review, this lens enables a broad description of the problem designers aim to address and how they achieve their objectives.

2.3 Methodology

The methodology for this scoping review is based on Arksey and O'Malley (2005) and with consideration to the recommendations made by Levac et al. (2010) and Daudt et al. (2013). To complete the review process, we followed the PRISMA extension for scoping reviews checklist (Tricco et al., 2018). The review comprised five key phases of the scoping process: (1) the research question identification, (2) relevant studies identification, (3) study selection, (4) data charting, and (5) summarizing and reporting the results. The optional "consultation exercise" is not conducted.

2.3.1 Research question

From initial motivation to review completion, our research question remained the same: “What is the current state of research on the use of neurophysiological measures and artificial intelligence techniques for mental state estimation in HCI within adaptive environments?” This research question and its motivation align strongly with the stated goals of a scoping review, where the process is “to map the literature on a particular topic or research area and provide an opportunity to identify key concepts; gaps in the research; and types and sources of evidence to inform practice, policymaking, and research” (Daudt et al., 2013).

2.3.2 Scoping process

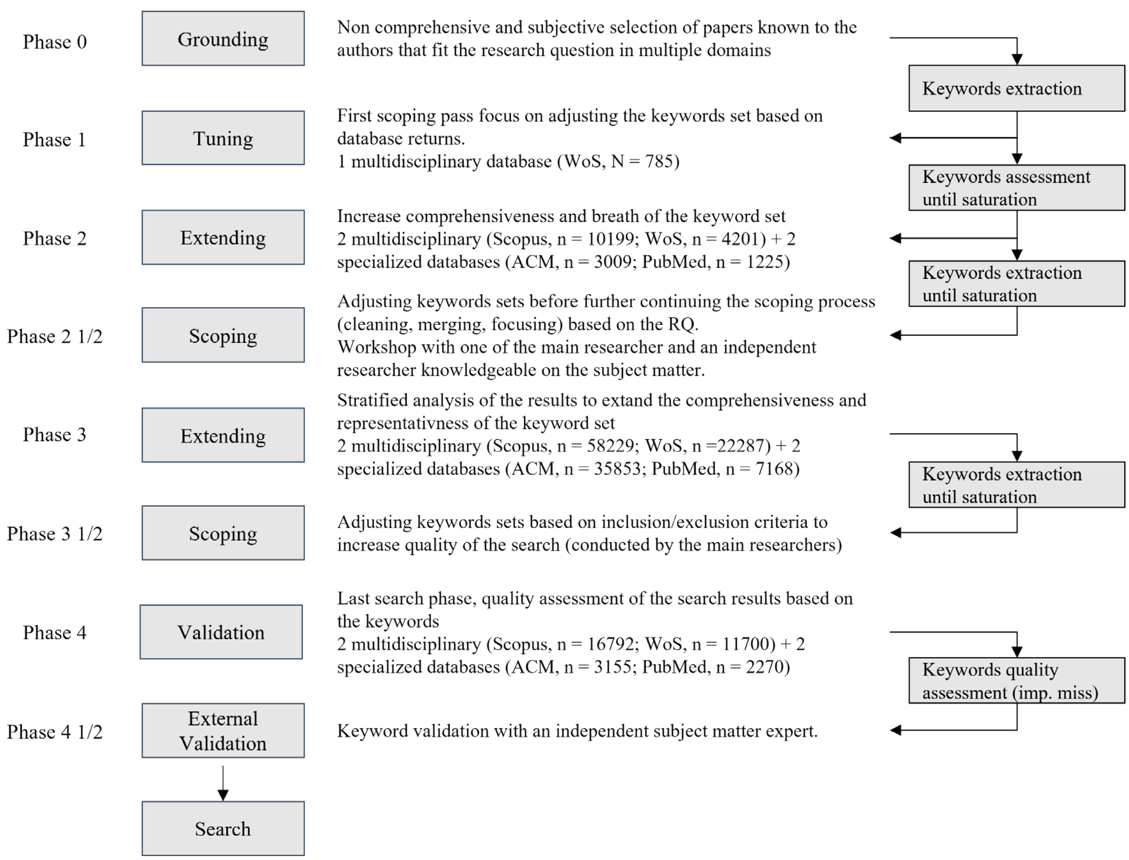
The scoping process was instigated by emphasizing keyword set creation. The multidisciplinary focus of the scoping review (e.g., neuroscience, engineering, medicine) required a rigorous survey of the terms used in each domain that refer to our research question. We divided the scoping process into five phases (0-4), 0) grounding, 1) tuning, 2-3) extending, and 4) validation. In addition, we utilized three collaborator workshops (phases 2 ½, 3 ½, 4 ½) (Figure 10). All queries, search dates, keywords, and data records were stored for each phase. The outcome of this scoping process was a set of queries for each database that has been refined during the multiple phases. The queries served during the search phase.

- Phase 0 – Grounding: The grounding phase consisted of listing a non-comprehensive set of keywords based on a set of papers we though were representative of the type of literature we wanted to capture to answer the R.Q.
- Phase 1 – Tuning: During the tuning phase, query design required multiple passes on Web of Science (WoS) to adjust the query for unwanted/unexpected results and to find the first balance between comprehensiveness and breath. After the fourth pass, an overview showed a majority of the pertinent papers for the R.Q. In total, 785 manuscripts resulting from the search query were exported with Title/Keywords/Abstract. We screened that information to expand our keyword list, thus increasing the depth and breadth of the scoping study. No judgment on keywords or quality was made; the objective was to capture the concepts used in different domains, and only those relevant to the R.Q. were assessed based on title and abstract. Keywords were assessed until saturation. After 237 papers (≈ 30%), the saturation point was reached.
- Phase 2 – Extending: Based on the expanded keyword set, we continued the scoping process on four databases representative of the targeted literature (i.e., multidisciplinary databases with WoS and Scopus, medicine with PubMed, and ACM for engineering). The same saturation strategy was used. The following breakdown represents the number of papers screened for their keywords per source: WoS = 131, Scopus = 185, ACM = 115, PubMed = 158.
- Phase 2 ½ – Scoping: The scoping phase involved a workshop conducted by the main research and a researcher knowledgeable of the subject that did not take part in the previous phases. The workshop aimed to refine the keyword sets by merging, cleaning, and adjusting keywords based on the R.Q.
- Phase 3 – Extending: The same methodology was applied for phase 3 as was used in phase 2. However, due to the significant number of records, a stratified selection of records was applied (first 10 every 100 records for the first thousand, then the first 10 every 1000 records after, with a maximum of 200 records screened). The goal was to ensure that no relevant literature was missed because of the number of results and to control for the ordering of the different databases.

- Phase 3 ½ – Scoping: Similar process to phase 2 ½, however, we adjusted the inclusion/exclusion criteria based on the results of the previous phase queries to increase the quality of the search.
- Phase 4 – Validation: The validation phase consisted of a subjective assessment of the search results based on the refined queries.
- Phase 4 ½ – External Validation: Similar process to phase 2 ½ and 3 ½. The workshop was conducted with an independent subject matter expert.

Figure 10

Scoping Strategy



2.3.3 Data sources and search strategy

The search phase was conducted on March 21, 2022, in seven electronic databases that span a comprehensive range of disciplines: multidisciplinary (Scopus, Web of Science, EBSCO), medicine (PubMed), business and economics (ProQuest ABI/INFORM Collection), psychology (PsycInfo), engineering (IEEExplore). IEEExplore was queried on March 22, 2022. Limits on dates (>= 2012 to today), language (English), and peer-reviewed records were used if the function existed.

The search string consisted of four categories of keywords which were gathered explicitly during the scoping process to cover the research question. The categories follow the subsequent framework: context of use AND (State AND Artifact Objectives) AND A.I. Within each category, the search operator OR was used between keywords. Context of use represents the relevant keywords to the environment for which the technological artifact was created. State includes mental state synonyms and concepts found in the literature. Artifact Objectives correspond to keywords covering the primary function of interest of the artifact, i.e., its adaptivity. A.I. corresponds to a broad selection of search terms for machine learning and predictive modeling. A first query was developed for WoS and then adapted to fit the format of each database searched. All queries and keywords are documented and available.

2.3.4 Eligibility criteria

We use the following inclusion criteria to screen the records: primary literature (peer-reviewed) such as journal or conference articles, (neuro)physiological measurements as input of the artifact (e.g., facial emotion recognition does not qualify), subjects are healthy adults population, operators the task and the neuro-adaptive artifact (e.g., worker, surgeon, air traffic controller, pilot), direct interaction between the operator and the adaptive artifact, and finally, the produced artifact is empirically tested. Moreover, manuscripts should be in either English or French to be assessed by the review team. As exclusion criteria, all records before 2012, reviews, editorials, and proceeding summaries are discarded. We exclude neuro-adaptive artifacts for therapy, rehabilitation, medication, artifact-patient interaction, and BCI for control as the study focuses on passive and reactive designs. No quality appraisal is performed.

2.3.5 Screening process

Covidence software (Veritas Health Innovation, 2023) supported the screening and data charting process. Two reviewers performed screening for record inclusion on title/abstract. Due to the number of records, each one is screened by one reviewer only, and to address this limitation, a training meeting was organized with the reviewers. Then, 50 records were randomly selected and screened independently by the reviewers, resulting in an inter-rater reliability of 0.88. Then, another meeting was conducted to discuss each record to ensure a common understanding of the procedure. Reviewers continued to meet during this stage to discuss progress, challenges, and ambiguities in the study selection.

Finally, four reviewers conducted full-text screening of the remaining records (reviewer 1: n = 427, reviewer 2: n = 366, reviewer 3: n = 70, reviewer 4: n = 49), and conflicts regarding the eligibility were resolved by consensus between the two leading reviewers. Two reviewers minimally screened each paper.

During the data charting process, if multiple manuscripts were based on the same experiment, we applied the following decision process: (1) if a journal article is part of the duplicates (e.g., journal and conference articles), we code only the journal article, (2) if published only through conference proceedings, the most recent one is selected. The purpose is to avoid artificial literature inflation on a specific domain/artifact/BCI in the results. This case was applied for 3 papers during the process, making 36 articles eligible to be included in the results.

2.3.6 Data charting

After the full text screening process, all relevant manuscripts are coded based on the form created by the authors. The form is developed based on the conceptual framework presented in the relevant section. Following (Levac et al., 2010) recommendations, charting is implemented in an iterative process during which reviewers update the conceptual model and the form. Descriptive numerical and qualitative thematic analysis will be performed on the data to answer the research question.

Data is charted following a conceptual framework derived from DSR presented in the "analytical framework" section. Data is collated and summarized considering three spaces that influence the resulting artifact: (1) the problem space, (2) the knowledge space, and (3) the solution space (Figure 1). The level of detail reported varied noticeably across the reviewed manuscripts. As the literature is heterogeneous regarding the expected presence of components relative to the spaces, we code the relevant information only if clearly stated. We first assessed the presence of the components before coding the excerpts. If the component (e.g., problem statement, research goals/questions) were not clearly written, it was labeled as "unclear/inferred." The extraction template is available in the appendix.

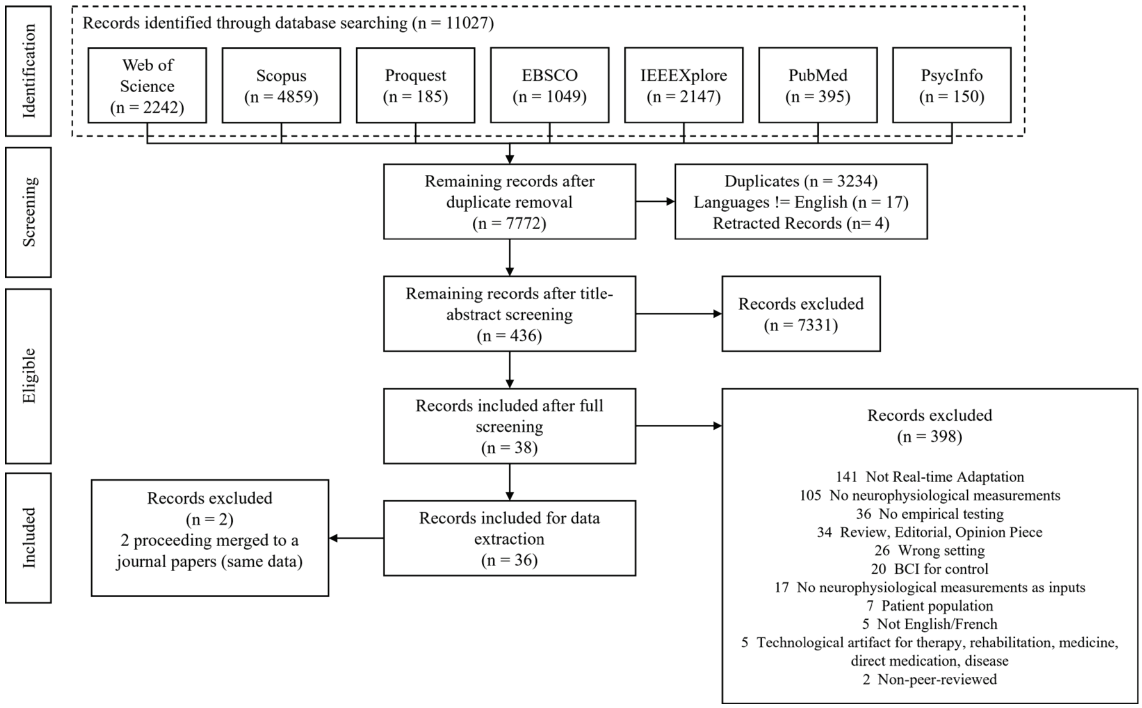
2.4 Results

2.4.1 Literature search and screening

The initial search phase, conducted between March 21 and 22, 2022, yielded 11,027 papers. Following the de-duplication procedure outlined by Bramer et al. (2016), 7,772 papers were screened based on their abstracts and titles, of which 436 met the eligibility criteria. Subsequent full-text screening resulted in the inclusion of 38 papers for data extraction, while 398 records were excluded (Figure 11). Figure 3 provides descriptive statistics regarding the reasons for exclusion. Primary causes for exclusion included a lack of real-time adaptation in the artifact (e.g., classification method or offline testing of adaptation) and the absence of neurophysiological measurements as inputs for the adaptive system (e.g., facial emotion recognition). Among the included manuscripts for data extraction, two were consolidated due to their reliance on the same experiments and data, resulting in a total of 36 papers for the extraction and charting phase.

Figure 11

PRISMA flow chart



Note. Details the flow of screened manuscripts during the different phases and the reasons for exclusion.

2.4.2 Study characteristics

The 36 manuscripts were spread between 2012 and 2022. The number of publications has been steadily increasing, with a range of one to four papers per year between 2012 and 2017 and five to eight papers per year from 2018 onwards (Figure 12). This trend is accompanied by a growing number of journal articles, indicating increased acceptance in the field and illustrating its ongoing emergence. Conference manuscripts and journal articles account for 47.22% and 52.78% of the publications, respectively. Most of the research was conducted in the United States (27.78%), followed by Italy (16.67%) and Australia (8.33%), Canada (8.33%), and Germany (8.33%) (see Table 6)

Figure 12

Publication type and frequency of publications per year

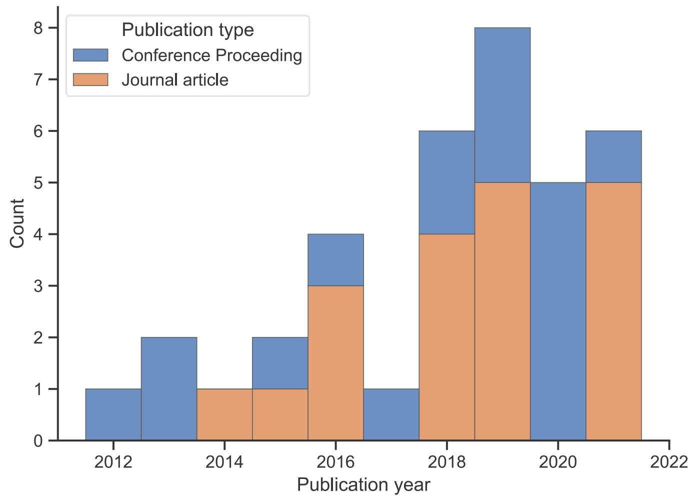


Table 6

Publication outlets and country

Study Characteristics (n = 36)		Count	(%)
Publication outlet	Conference Proceeding	17	47.22%
	Journal article	19	52.78%
Country	Australia	3	8.33%
	Canada	3	8.33%
	China	2	5.56%
	Denmark	1	2.78%
	France	2	5.56%
	Germany	3	8.33%
	Italy	6	16.67%
	Netherlands	1	2.78%
	Poland	1	2.78%
	Portugal	1	2.78%
	Taiwan	1	2.78%
	U.K.	2	5.56%
	United States	10	27.78%

2.4.3 Problem space

In the manuscripts, the problem space is characterized by the environments and the problematization. On one hand, the environment represents the socio-technical system from which the problem emerges. On the other hand, problematization denotes the researchers' interpretation of the problem they are attempting to solve with a neuro-adaptive artifact.

2.4.3.1 Environment

The included 36 manuscripts are spread across 10 different domains. Aeronautics is the domain with the largest number of occurrences, with 7 observations representing 19.44% of the research charted. The training domain, with 7 occurrences, represents the same percentage as aeronautics. Those papers leverage neuro-adaptive systems for the training of a skill (e.g., stress management) (El-Samahy et al., 2015), and attention regulation (Zargari Marandi et al., 2019). We also found manuscripts in business (4, 11.11%), education (3, 8.33%), robotics (3), transport (2, 5.56%), video games (2), and art (2). The smallest occurrences belong to the smart home (1, 2.78%) domain and virtual agent design (1), each with only 1 observation, representing 2.78% of the total. The domain could not be labelled for 4 manuscripts (11.11%).

The defined target tasks were quite diverse in the included research. Neuro-adaptive systems were created to support hypothetical surveillance tasks in 7 studies (19.44%). We found 3 (8.33%) manuscripts for both human-robot and learning tasks. Target driving (2, 5.56%), game (2), self-regulation (2), and supervision tasks (2) were described to motivate the creation and integration of the created artifact. Tasks such as cognitive, computer, concentration, cooperation, information seeking, listening, natural, relaxation, and training tasks showed one (1, 2.78%) occurrence. Finally, 6 (16.67%) manuscripts were unclear on the target task. In this case, the manuscripts do not refer to a specific task when motivating the creation of the neuro-adaptive systems. For example, El-Samahy et al. (2015) describe the target users and the cause of the problem but refer to the task in vague terms. The manuscripts refer to the "chronic mental stress" caused by "job environments" in "human's work" (El-Samahy et al., 2015, p. 1). Another manuscript documents the design of a neuro-adaptive virtual character but does not refer to a hypothetical task of applications when motivating the study (Aranyi et al., 2016).

In most cases, the target users for the design of a neuro-adaptive system were unclear/inferred (15, 41.67%). Artifacts built for air traffic controllers (4, 11.11%), learners (4), and office workers (4) showed the most occurrences. They are followed by unmanned vehicle operators (3, 8.33%), video game players (2, 5.56%), drivers (2, 5.56), machinery operators (1, 2.78%), and production line workers (1).

2.4.3.2 Problematization

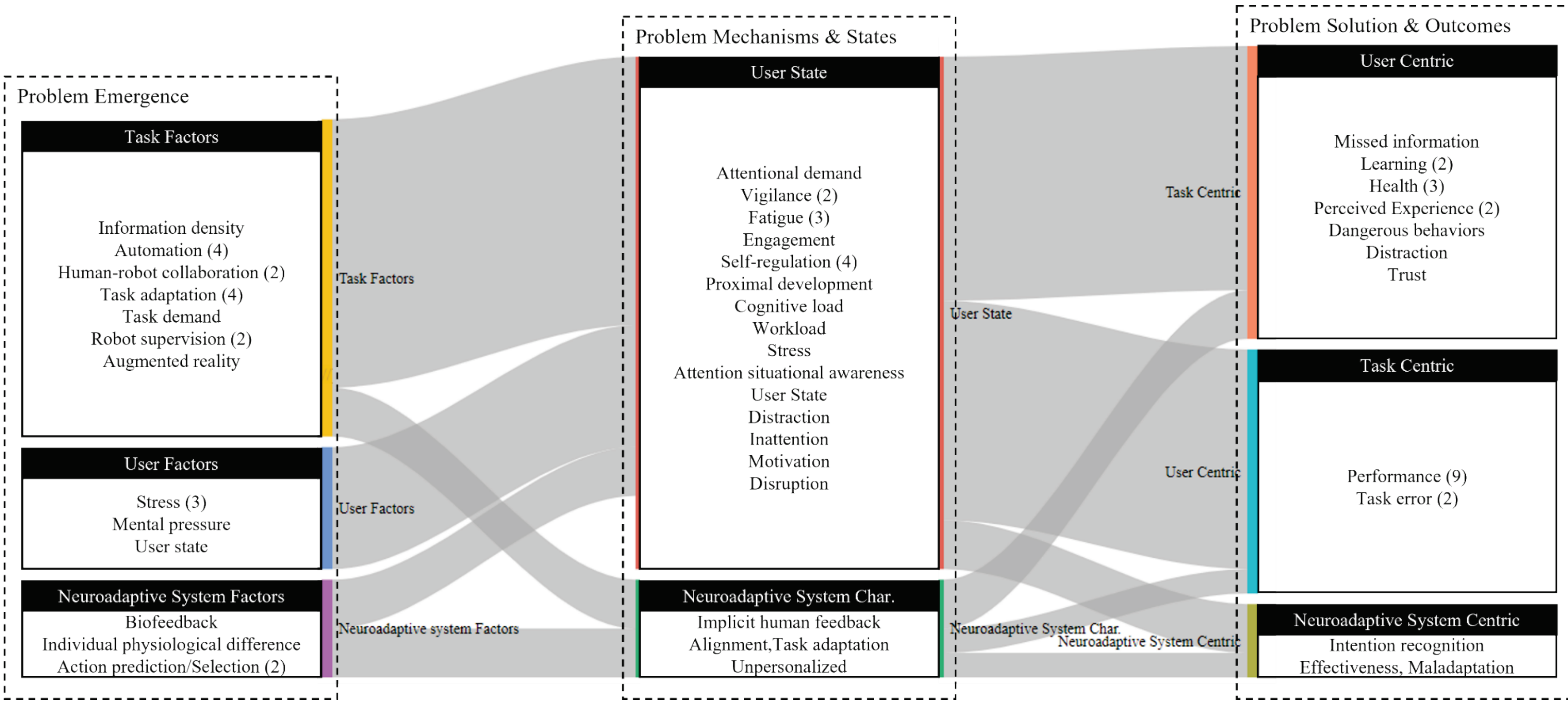
Problematization represents the research objective/question, the artifact goals, and the problem's relevance, as stated in the research. Firstly, three (8.33%) manuscripts clearly stated a research question, while 33 (91.67%) did not. In those cases, the research question was unclear or inferred. However, 21 (58.33%) manuscripts presented clear research objectives.

Secondly, the artifact objective was clearly defined in 22 papers (61.11%). For the remaining 14, the objective was either unclear (11, 30.56%) or merged with the research objectives (3, 8.33%). When clearly articulated, the artifact objectives were expressed as a general objective (e.g., "the artifact should do...") in 13 studies (36.11%), as hypotheses in 4 manuscripts (11.11%), and as a set of requirements in 5 instances (13.89%).

Thirdly, to understand how the problem's relevance is motivated, we qualitatively analyzed the problem statements. The problem statement describes the problem that the neuro-adaptive system aims to resolve. It was clearly reported in 24 papers (66.67%) of the manuscripts. On the contrary, it was unclear in 12 (33.33%) of them. We divided the problem states into three components (refer to Figure 13 for a visual representation in a Sankey diagram): the factors that contribute to the emergence of the problem, the mechanisms and states impacted by these factors, and, ultimately, the outcomes resulting from these. The Sankey diagram (Figure 13) provides a qualitative representation of how the problem statements are constructed to justify the need for neuro-adaptive artifacts.

Figure 13

Sankey diagram of problematization flow



Note. Problem statements were grouped into grand categories depicting the emergence of the problem, the mechanisms, and the outcomes. Nodes represent categories within these grand categories. The size of the links represents the logical flow between these categories in the manuscripts.

We found that they emerged from the task (e.g., information density (Toreini et al., 2020), automation (Demazure et al., 2019), system supervision (Breslow et al., 2014)), the user (e.g., stress (Azgomi et al., 2021; Parnandi & Gutierrez-Osuna, 2021), mental pressure (Zhang et al., 2021)), or the neuro-adaptive system design (e.g., biofeedback design (Raaijmakers et al., 2013), individual physiological differences (Labonte-Lemoyne et al., 2018)). These factors lead to problematic situations, such as unwanted user states (e.g., fatigue (Peternel et al., 2018), distraction (Pavlidis et al., 2021), workload (Breslow et al., 2014)) or design challenges (e.g., encoding implicit human feedback (Kim et al., 2020), or alignment between the adaptation and the user expectation (Causse et al., 2019; Govindarajan et al., 2018)). Finally, while problematizing, we found that manuscripts define the outcome either as user centric (e.g., learning, perceived experience, health), task centric (e.g., performance or error at the task), or system centric (e.g., effectiveness, maladaptation).

2.4.4 Knowledge space

Neuro-adaptive systems research is grounded upon prior literature concerning analogous artifact designs, theoretical frameworks, and the characterization of neurophysiological patterns. We have categorized these elements within our analytical framework under the knowledge space. This knowledge space encompasses descriptive knowledge (e.g., patterns and theories) and prescriptive knowledge (e.g., instantiations and methodologies). Table 7 showcases the knowledge utilized in the examined manuscripts to construct the artifacts, with components coded only if they were clearly stated.

2.4.4.1 Descriptive knowledge

Manuscripts utilized descriptive knowledge from the relevant literature to characterize the neurophysiological patterns upon which their neuroadaptive systems are based. Brain patterns such as Event Related Potential (ERP) or oscillatory activity were depicted in 8 studies. For instance, Kim et al. (2020) and Schiatti et al. (2018) built upon Error-Related Potentials, the neurophysiological responses associated with human error processing. Other studies described pertinent oscillatory activities under specific tasks or mental states (Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016; Dey et al., 2019; Di Flumeri et al., 2019; Labonte-Lemoyne et al., 2018; Vortmann & Putze, 2020).

Gaze patterns were examined in three cases for attention (Causse et al., 2019; Toreini et al., 2020) and fatigue (Zargari Marandi et al., 2019). Physiological patterns were characterized in four instances for stress (Azgomi et al., 2021; Raaijmakers et al., 2013) and arousal (Pavlidis et al., 2021). Theories were employed to guide the design of neuro-adaptive systems in 6 manuscripts. It is worth noting that other manuscripts utilized concepts and explanations from these theories, particularly cognitive load theories, but did not explicitly state their use. We observed a relatively small proportion of manuscripts that thoroughly describe and explain the neurophysiological processes and patterns. As the subsequent prescriptive knowledge section demonstrates, most manuscripts employ measurement methods for these neurophysiological patterns.

2.4.4.2 Prescriptive knowledge

The most prevalent type of prescriptive knowledge utilized in the examined manuscripts involved a descriptive review of relevant instances of similar or closely related neuro-adaptive systems. A total of 28 (77.785) manuscripts cited previous artifacts to inform or justify the design and evaluation of the current neuro-adaptive systems under evaluation. For instance, some manuscripts referenced general BCI literature (Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016; Ramos et al., 2021; Tseng et al., 2012) specific instances such as BCI in Augmented Reality (AR) (Dey et al., 2019; Vortmann & Putze, 2020), or biofeedback (Azgomi et al., 2021; Parnandi & Gutierrez-Osuna, 2021; Wang et al., 2019; Zargari Marandi et al., 2019; Zhang et al., 2021). Moreover, manuscripts drew upon specific artifacts from outside the neuro-adaptive systems literature, such as Causse et al. (2019) incorporating Decision Support System and Case-based reasoning systems, Peternel et al. (2018) focusing on human-robot collaboration designs, Govindarajan et al. (2018) referencing the design of Advanced Driver Assistance Systems (ADAS), or Yuksel et al. (2016) anchoring neuro-adaptive systems within Computer-based Education and Intelligent Tutoring Systems.

Table 7

Knowledge space (descriptive and prescriptive knowledge)

Knowledge space component	Example	Count	%
Descriptive Knowledge			
Brain patterns	ERP (2), Oscillatory Activity of user mental states (5), Hemodynamic Response (1)	8	22.22%
Gaze patterns	Attention (2), fatigue (1)	3	8.33%
Physiological patterns	Arousal (1), stress (2), physiological linkage (1)	4	11.11%
Psychological theory	Arousal theory (1), automatic control theory (1), cognitive load theory (2), motivation theory (1), reinforcement theory (1)	6	16.67%
Prescriptive Knowledge			
Defined constructs	Fan-out, internal/external/visual attention, mental workload, vigilance	10	27.78%
Psychophysiological inference	Mental workload – brain patterns, attention – brain patterns, error processing – ERP, stress – physiological patterns	22	61.11%
Models	Case-base reasoning, closed loop approach, Fan-out model, Dynamic Difficulty Adjustment framework	6	16.66%
Methods	Measurements (20) (e.g., classification algorithms, signal transformation techniques, indexes), practice (2) (e.g., EEG measurement in AR, transcranial direct current stimulation), Task (e.g., stressors, paradigm)	24	66.67%
Design theory	Design science research (2)	2	5.56%
Instantiations	Affective Systems, Adaptive Systems, related BCI, Biofeedback and Neurofeedback systems, Computer based education, Intelligent tutoring systems, Case-based reasoning in DSS	28	77.78%

Manuscripts discussed relevant methods in 24 instances (66.67%). In most cases, the papers referred to measurement approaches such as classification algorithms, signal transformation techniques, and indexes. For example, some manuscripts cited algorithms used for classifying specific mental states based on sensor data (Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016; Azgomi et al., 2021; Kim et al., 2020; Ramos et al., 2021; Vortmann & Putze, 2020). Other manuscripts reviewed the utilization of instruments for measuring targeted states (e.g., EEG, GSR) (Demazure et al., 2019; El-Samahy et al., 2015; Szafir & Mutlu, 2012). Some manuscripts referred to specific practices, for example, Dey et al. (2019) reviewed the use of EEG in AR. Additionally, several manuscripts examined tasks designed to elicit the targeted user states (El-Samahy et al., 2015; Trachel et al., 2018; Zhou et al., 2015).

In 22 manuscripts (61.11%), evidence was presented to support the psychophysiological inferences underlying the neuro-adaptive systems. The employed strategy involved listing empirical evidence that substantiated the inferences made. For example, Aranyi et al. (2016) described the correlates of the affective-motivational dimension of approach in the brain, referring to changes in oxygenated and deoxygenated hemoglobin measured with fNIRS. Breslow et al. (2014) discussed the link between fixation, attention allocation, and decision-making. Manuscripts also referred to oscillation activities correlated with engagement (Chaouachi et al., 2015), mental workload (Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016; Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016; Dey et al., 2019), or vigilance (Di Flumeri et al., 2019).

Existing models, such as frameworks and mathematical models, were utilized in 6 manuscripts. For example, (Breslow et al., 2014) built upon an established mathematical model to compute the number of autonomous aircraft operators that could be managed. In their research, they adapted this model to include the operator state. Labonte-Lemoyne et al. (2018) informed their neuro-adaptive systems with the “dynamic difficulty adjustment framework.” Additionally, we identified 2 papers (5.56%) employing a design theory (Demazure et al., 2019; Toreini et al., 2020), both utilized a Design Science Research approach.

2.4.5 Solution space

2.4.5.1 Design and development, evaluation approaches and methods

The solution space categories and components are reported in Table 8. To map how the manuscripts were positioned, we subjectively evaluated the contribution based on how the authors motivated the paper and how descriptive/prescriptive knowledge was leveraged. Half of the manuscripts were positioned as exaptation (50%), followed by improvement (25%), and routine design (22.22%). Those results show that most papers either extend known solutions to a novel problem (exaptation) or develop novel solutions to a known problem (improvement).

Most of the included manuscripts used laboratory experiments with 34 studies (94.44%) to evaluate the designed neuro-adaptive system. Only one manuscript leveraged simulation techniques to study the prototype (Azgomi et al., 2021), and one evaluated their artifacts during an uncontrolled field study (Ghandi et al., 2021). 38.89% of the studies had student participants during the artifact's evaluation. Those studies did not specify specific characteristics that would be congruent with the task. However, studies in aeronautics focused on neuro-adaptive artifacts for air traffic controllers. All of them either used aeronautics students (5.56%) or professional air traffic controllers (8.33%). A significant proportion of the manuscripts did not clearly state the participant sample used (44.44%). Moreover, the average number of participants was 14.78 (Std = 10.29), with a minimum number of 1 and a maximum of 47 participants.

The majority of the assessed research used experimental design evaluation methods (97.22%), while only one used a descriptive evaluation method (2.78%) to assess the neuro-adaptive system. 38.89% of the papers used an experimental evaluation only. However, several studies combined multiple evaluation methods (58.33%). Analytical approaches (e.g., dynamic analysis and architecture analysis) were the most popular, with 27.78% of the papers combining experimental evaluation. We observed a heterogeneous use of methods in the sample of manuscripts, such as using descriptive (i.e., informed arguments) or observational methods with 8 and 6 studies, respectively.

Table 8

Solution space component characteristics

Solution space component	Count	%
Design and Development		
Exaptation	18	50.00%
Improvement	9	25.00%
Invention	1	2.78%
Routine design	8	22.22%
Design Evaluation approach		
Field experiment	1	2.78%
Laboratory experiment	34	94.44%
Simulation	1	2.78%
Actual subjects		
Aeronautic Students	2	5.56%
Air Traffic Controllers	3	8.33%
Psychology Students	1	2.78%
Students	14	38.89%
Unclear/inferred	16	44.44%
Design Evaluation Method		
Descriptive	1	2.78%
Experimental	14	38.89%
Experimental / Analytical	10	27.78%
Experimental / Descriptive	3	8.33%
Experimental / Observational	1	2.78%
Experimental / Analytical / Descriptive	2	5.56%
Experimental / Analytical / Observational	3	8.33%
Experimental / Observational / Descriptive	2	5.56%
Tasks		
Ecologically Valid Paradigm	33	91.67%
Validated/Standard Paradigm	3	8.33%
Number of participants*	14.78 participants (Std = 10.29 [min = 1, max = 47])	

Note. * In the case of multiple studies in one research, we compiled the participants of the evaluation study of the adaptive artifacts

2.4.5.2 Functional components

Neuro-adaptive systems are composed of functional components. In this section, we map the sensors, the feature extractions (i.e., transforming the signal into features representing the neurophysiological mechanism targeted), and the features translator (i.e., translating the features into labels/continuous values).

Table 9

Sensors as inputs of the neuro-adaptive system

Input	Sensor	Count	%
Instrument			
	Electroencephalography	14	38.89%
	Galvanic skin response	4	11.11%
	Oculometry	4	11.11%
	Functional near-infrared spectroscopy	2	5.56%
	Breathing sensors	1	2.78%
	Electrocardiography	1	2.78%
	Electromyography	1	2.78%
	Thermal facial imaging	1	2.78%
Fusion			
	Fusion (ECG, GSR)	1	2.78%
	Fusion (ECG, GSR, EMG)	1	2.78%
	Fusion (ECG, OCU)	2	5.56%
	Fusion (EEG, GSR, BVP)	1	2.78%
	Fusion (EEG, OCU)	2	5.56%
	Fusion (EEG, TFI)	1	2.78%

As reported in Table 9, the primary neuro-adaptive systems' input sensors in the included manuscripts were EEG alone with 14 papers (38.89%), followed by GSR and oculometry with four papers (11.11%). In descending order, two studies used FNIRS (5.56%) alone, followed by breathing sensors (1, 2.78%), electrocardiography (1), electromyography (1), and thermal facial imaging (1). Sensor fusion was employed in

22.22% of the research. In this case, the configurations of instruments are heterogeneous.

Table 10 displays the different sensors, domain (i.e., time or frequency) of features extracted from the signal, and features employed as inputs of the feature extractors. As we show later, overlaps might occur as some studies generated multiple features in different domains to create feature vectors. EEG is the most popular sensor. It presents a majority of signal transformation from the time-domain to the frequency-domain with 18 papers. The research uses features derived from established signal transformation, such as Fast Fourier Analysis in the frequency domain. Another layer of transformation is applied to derive the band power of specific ranges linked to psychophysiological inferences (e.g., alpha, beta, theta, gamma). We will see in the next section that those features serve either in a features vector, often for the use of machine learning (e.g., (Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016; Vortmann & Putze, 2020)), or to construct indexes (e.g., (Demazure et al., 2019; Zhang et al., 2021)).

The research mainly derives features in the time-domain (6) for GSR by characterizing the signal. Similar observations are made for the oculometry (8). However, in this case, features can be further categorized: (1) eye movements features (e.g., dwell time (Causse et al., 2019; Toreini et al., 2020), fixations (Breslow et al., 2014), saccades (Zargari Marandi et al., 2019), or gaze entropy (Lim et al., 2021)), (2) pupillometry features (e.g., pupil dilatation diameter (El-Samahy et al., 2015)) and blink movements features (e.g., blinks frequency (Zargari Marandi et al., 2019)).

Studies with neuro-adaptive systems with ECG as sensor inputs implemented transformation in both the time domain (5) and frequency domain (2). In the time domain, the features employed were heart rate (Nalepa et al., 2019), inter-beats interval (Karthikeyan & Mehta, 2020), and descriptive statistics to characterize the signal further. In the frequency domain, spectral powers in low/high frequency ranges were computed to serve as features directly or create ratios (Darzi & Novak, 2021; Karthikeyan & Mehta, 2020).

For the rest of the sensors, signals were mainly transformed into features in the time domain. For example, the relative change in oxy-HB or left/right asymmetry was computed for fNIRS (2) (Aranyi et al., 2016; Yuksel et al., 2016). For electromyography (n = 2), the signal is derived into descriptive statistics like root-mean-square to estimate facial muscular activity (Darzi & Novak, 2021). In another case, when applied to the shoulder, the signal was transformed into features such as muscle activation level and maximal voluntary contraction (Peternel et al., 2018). Finally, breathing rate and its variation were used with breathing sensors (n = 2) (Darzi & Novak, 2021; Parnandi & Gutierrez-Osuna, 2021)

Table 10

Sensors, features extractions mapping (sensors fusions are split across the individual sensors)

Sensors	Domain	Features	Reference
Electroencephalography (n = 18)	Time-domain (n = 2)	XDAWN, ERPs characteristics (descriptive statistics)	(Kim et al., 2020; Schiatti et al., 2018)
	Frequency-domain (n = 17)	Frequency bands with Fast Fourier Analysis FFT (Alpha, Beta, Theta, Gamma), band ratio (theta/beta), relative power band, individual alpha frequency, sensorimotor rhythms, alpha asymmetry	(Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016; Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016; Chaouachi et al., 2015; Demazure et al., 2018; Dey et al., 2019; Di Flumeri et al., 2019; Ghandi et al., 2021; Govindarajan et al., 2018; Labonte-Lemoyne et al., 2018; Ramos et al., 2021; Schiatti et al., 2018; Szafir & Mutlu, 2012; Trachel et al., 2018; Tseng et al., 2012; Vortmann & Putze, 2020; Wang et al., 2019; Zhang et al., 2021)
Galvanic skin response (n = 7)	Time-domain (n = 6)	Signal characteristics (peaks, amplitude of peaks, durations, rising time, Mean skin conductance, skin conductance difference, skin conductance response frequency, mean skin conductance response amplitude, standard deviation of response amplitude, binarized pulse)	(Azgomi et al., 2021; Darzi & Novak, 2021; Larradet et al., 2017; Nalepa et al., 2019; Raaijmakers et al., 2013; Zhou et al., 2015)
	Unclear/Inferred (n =1)		(Ghandi et al., 2021)
Oculometry (n = 8)	Time-domain (n = 8)	Dwell time, Percentage of the duration of closed eyes to opened eyes, Saccade Frequency, Saccade Peak Velocity Amplitude Relationship, Pupil Diameter Interquartile Range, Gaze	(Breslow et al., 2014; Causse et al., 2019; Di Flumeri et al., 2019; El-Samahy et al., 2015; Lim et al., 2021; Toreini et al., 2020; Vortmann & Putze, 2020; Zargari Marandi et al., 2019)

		entropy Frequency of blinks, saccades, fixations Number of blinks, saccades, fixations Mean duration of blinks, fixations, saccades Pupil diameter mean, coefficient of variation, interquartile range, mean value of their peak velocity, amplitude, curvature Peak amplitude of saccadic acceleration and deceleration profiles, intersaccadic intervals	
Functional near-infrared spectroscopy (n = 2)	Time-domain (n = 2)	Optical density, relative change in oxy-HB, mean/linear regression slope, Left-Right Asymmetry	(Aranyi et al., 2016; Yuksel et al., 2016)
Breathing sensors (n = 2)	Time-domain (n = 2)	Breathing rate, rate variation, standard breathing rate	(Darzi & Novak, 2021; Parnandi & Gutierrez-Osuna, 2021)
Electrocardiography (n = 5)	Time-domain (n = 5)	Inter-beat Interval (IBI), Heart Rate, median of absolute deviation from the average IBI, standard deviation of intervals between consecutive beats (C.B.), root mean square of successive differences between C.B. intervals, standard deviation of successive differences between C.B. intervals, proportion of differences between C.B. intervals, peaks, amplitude of peaks	(Darzi & Novak, 2021; El-Samahy et al., 2015; Karthikeyan & Mehta, 2020; Lim et al., 2021; Nalepa et al., 2019)
	Frequency-domain (n = 2)	Spectral powers across frequency regimes, power of low frequencies (L.F.), power of high frequencies (H.F.), ratio of LF/HF	(Karthikeyan & Mehta, 2020)
Electromyography (n = 2)	Time-domain (n = 2)	FEMG : root-mean-square (RMS), maximum and minimum RMS	(Darzi & Novak, 2021; Peternel et al., 2018)

		Shoulder EMG : maximal voluntary contraction, muscle activation level	
Thermal facial imaging (n = 2)	Frequency-domain (n = 2)	Mean perinasal perspiration, mean temperature of region of interest (i.e., forehead, left eye, right eye, nose)	(Govindarajan et al., 2018; Pavlidis et al., 2021)
Blood volume pressure (n = 1)	Unclear/Inferred (n =1)		(Ghandi et al., 2021)

Table 11 showcases how the included research uses these features to construct index or feature vectors. As previously stated, techniques like Fast Fourier Transform or bandpass filtering are used to extract the absolute or relative powers in EEG signals. For a neuro-adaptive system, this process represents the end of the feature extraction phase, and the created data serve as inputs for the features translator. We grouped the literature into two primary techniques: a compilation of features to create a features vector (20, 55.56%), or an aggregation of features to compute indexes/scores or relative indexes (15, 41.67%). A features vector represents an n-dimensional vector of features that often serves as a machine learning model. In the included manuscript, a compilation of features was used when a fusion of sensors was employed (6). It was also used for EEG (6), oculometry (3), GSR (2), and ECG (1).

In the case of aggregation techniques, two approaches were identified. Studies derived indexes or scores (5, 13.89%) from the features or built relative indexes (10. 27.78%) by creating an index and comparing it to a baseline. Put differently, an index is an aggregated value created from chosen features. Conversely, a relative index represents an aggregated value that not only stems from selected features but is also derived from comparisons with the same index computed from the user in a different cognitive/physiological state (e.g., baseline, moving average). However, using indexes and feature vectors are not mutually exclusive. For example, Chaouachi et al. (2015) created an index based on the Beta/(Theta + Alpha) ratio and used it in a feature vector to train a machine learning algorithm.

Table 11

Feature-domain, features extractor, and sensors

Features Extractor	Sensors	F.D.	TD	TD + FD	Total	%
Features Compilation						
Features vectors		6	5	9	20	55.56%
	Fusion	3	2	3	8	22.22%
	EEG	3		3	6	16.67%
	OCU		2	1	3	8.33%
	GSR		1	1	2	5.56%
	ECG			1	1	2.78%
Features Aggregation						
Index/Score		5			5	13.89%
	EEG	3			3	8.33%
	GSR	1			1	2.78%
	TFI	1			1	2.78%
Relative index		5	5		10	27.78%
	EEG	5			5	13.89%
	FNIRS		2		2	5.56%
	BS		1		1	2.78%
	EMG		1		1	2.78%
	GSR		1		1	2.78%
No transformation			1		1	2.78%
	OCU		1		1	2.78%
Total		16	11	9	36	100.00%

Note. TD = Time-domain, FD = Frequency-domain

Table 12 shows how the feature translator outputs are utilized to derive logical control via the features translator. From the 20 manuscripts utilizing feature vectors, 16 (44.44%) built machine learning models to learn the relationship between the features and the outcome. Super Vector Machine (SVM) are popular with 6 occurrences (Schiatti et al., 2018; Yuksel et al., 2016), followed by Linear Discriminant Analysis (LDA) and some variations like asSWLDA (Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016; Di Flumeri et al., 2019). In most cases, machine learning classifies discrete outputs. For example, Schiatti et al. (2018) classify the presence of an error potential in EEG while supervising a robot's learning behaviors. In this case, the output is binary and represents the presence of the expected neurophysiological response. Similarly, Yuksel et al. (2016) learn from features extracted from fNIRS signal high and low cognitive workload. However, it is possible to derive continuous output using machine learning (Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016).

Moreover, the literature shows that a layer of fuzzy logic can be added on top of machine learning techniques. Fuzzy methods use condition/logic decisions to generate the final output. For example, Lim et al. (2021) classified the workload into three classes (low to high), attention (low to high), and performance (low to high). Then, they added a fuzzy layer that would translate combinations of those three variables into three automation levels.

Relative index and index/scores were mainly used with fuzzy methods, with 7 (19.44%) and 4 (11.11%) occurrences, respectively. In those cases, the feature translators are built based on thresholds. The output is derived from comparing the index during the task with the index during a baseline. For example, Labonte-Lemoyne et al. (2018) constructed fixed thresholds; when the real-time index is two times superior or inferior to the baseline, the system triggers an adaptation. Demazure et al. (2019) built a similar approach with dynamic thresholds derived from a calibration phase and the moving average during the task.

Table 12

Transformation and feature translators

Transformation	Feature Translator Category	Feature Translator	Count	%
Features vector			20	55.56%
	Fuzzy methods / Machine learning		4	11.11%
		Adaptive neuro-fuzzy inference systems (ANFIS) (1), Mamdani-type fuzzy model (1), Thresholds (1), Unclear/Inferred (1)		
	Machine learning		16	44.44%
		Gradient boosting (1), K-NN (2), LDA (1), Logistic regression (1), LSTM (1), RNN (1), Random Forest (2), Stepwise Linear Discriminant Analysis (asSWLDA) (3), SVM (6), Naïve Bayes (1)		
No transformation			1	2.78%
	No translator		1	2.78%
		No translator (1)		
Relative index			10	27.78%
	Fuzzy methods		7	19.44%
		Thresholds (6), Statistical Thresholds (1)		
	Machine learning		2	5.56%
		SVM (1), Gaussian Process Regression (1)		
	Unclear/Inferred		1	2.78%
		Unclear/Inferred (1)		
Index/Score			5	13.89%
	Fuzzy methods		4	11.11%
		Thresholds (3), Statistical Thresholds (1)		
	Unclear/Inferred		1	2.78%
		Unclear/Inferred (1)		

Table 13 presents the psychophysiological inference grouped by the high-level mechanisms targeted by the neuro-adaptive systems. Such artifacts are built on the assumption that changes in the extracted characteristics of the signal represent the emergence of a mental state. The literature shows that while the psychophysiological inferences are quite dispersed, they can be regrouped within high-level mechanisms. Neuro-adaptive systems designs centered on attentional mechanisms in 14 (38.88%) manuscripts. The studies focus on states like attention (3) (Lim et al., 2021; Wang et al., 2019; Zhang et al., 2021), engagement (2) (Chaouachi et al., 2015; Szafir & Mutlu, 2012) or visual attentional allocation (2) (Causse et al., 2019; Toreini et al., 2020). For example, Causse et al. (2019) built and evaluated a decision support system in air traffic control tasks for route suggestions on an airport tarmac. The neuro-adaptive system leverages users' attention allocation via fixations to weight parameters based on past gaze behaviors in route selection. The authors argue that the approach increases the system's performance and aligns the system's suggestions with the user decision-making process.

In 6 manuscripts, the artifacts target affective mechanisms (16.67%) referred as the user's affective state (2), valence/arousal (3), or emotion (1). For example, Darzi and Novak (2021) used ECG, GSR, and EMG to measure arousal and valence. They specify the relationship between the sensors, the features, and the psychophysiological inferences made. With ECG, they provide support that links interbeat intervals with tonic arousal. Mean conductance and conductance response frequency extracted from the GSR signal is related to arousal in the manuscript. The root mean square of the EMG signal at the zygomaticus major and the corrugator supercilia are influenced by smiling and frowning, respectively, and are correlated with the user's positive and negative affective states.

The third targeted mechanism is related to the user's information processing, with 5 manuscripts (13.89%). Referred to by different constructs (e.g., Cognitive load, Cognitive Workload, Mental workload, Overload), this research designs and builds neuro-adaptive mechanisms linked to the demand on working memory and information processing. Three manuscripts measure brain responses with EEG (2) and FNIRS (1), while the two others utilize oculometry (1) and GSR (1). For example, Zhou et al. (2015) derive time (i.e., peaks) and frequency domain (i.e., average power below 1 Hz) features from the GSR signal to classify the level of cognitive load with SVM, Naïve Bayes, and random forest.

Five manuscripts (13.89%) targeted stress mechanisms for their psychophysiological inferences. Two papers conceptualized it as cognitive stress (Azgomi et al., 2021; El-Samahy et al., 2015), while 2 others framed it as physiological stress (Parnandi & Gutierrez-Osuna, 2021; Raaijmakers et al., 2013). One study targeted relaxation to drive biofeedback (Larradet et al., 2017). Most studies use physiological sensors such as ECG, GSR, and BR. One study used ECG and OCU (El-Samahy et al., 2015). For example, El-Samahy et al. (2015) used feature vectors composed of heart rate variability and pupil dilatation features to predict task performance using a fuzzy clustering technique (i.e., Mamdani-type fuzzy model). Interestingly, Raaijmakers et al. (2013) validated self-regulation tasks with a third measurement instrument. ECG and GSR were used to measure physiological stress and drive the neuro-adaptive system to promote regulation in the user. In addition, the study used EEG to monitor the effect of stress regulation on change in frontal alpha asymmetry, which can be linked to cardiac activity modulation, per the authors.

Two manuscripts (5.56%) utilized fatigue mechanisms at both the cognitive (Zargari Marandi et al., 2019) and the physical (Peternel et al., 2018) levels with OCU and EMG, respectively. Zargari Marandi et al. (2019) developed a neuro-adaptive system that triggers biofeedback based on user fatigue during computer work. The artifact applies pupil diameter measurements to infer fatigue and prompt feedback to consider microbreaks. Peternel et al. (2018) used EMG to measure operators' physical fatigue to modulate task allocation during a collaborative and physically demanding task with a robot.

Table 13

Mechanisms and psychophysiological inference

Mechanisms Targeted	Psychophysiological Inference	Sensors (Count)	Count	%
Attentional Mechanisms			14	38.88%
	Alertness	EEG (1)	1	2.78%
	Arousal	TFI (1)	1	2.78%
	Attention	EEG (2), Fusion (ECG, OCU) (1)	3	8.33%
	Covert Visuospatial Attention	EEG (1)	1	2.78%
	Engagement	EEG (2)	2	5.56%
	Steady-State Visually Evoked Potential (SSVEP)	Fusion (EEG, OCU)	1	2.78%
	Sustained attention	EEG (1)	1	2.78%
	Vigilance	Fusion (EEG, OCU) (1), ECG (1)	2	5.56%
	Visual Attention Allocation	OCU (2)	2	5.56%
Affective Mechanisms			6	16.67%
	Affective state	FNIRS (1), Fusion (EEG, TFI) (1)	2	5.56%
	Emotion	Fusion (EEG, GSR, BVP) (1)	1	2.78%
	Valence/Arousal	EEG (1), Fusion (ECG, GSR) (1), Fusion (ECG, GSR, EMG) (1)	3	8.33%
Information Processing Mechanisms			5	13.89%
	Cognitive load	GSR (1)	1	2.78%
	Cognitive Workload	FNIRS (1)	1	2.78%
	Mental workload	EEG (2)	2	5.56%
	Overload	Oculometry (1)	1	2.78%
Stress Mechanisms			5	13.89%
	Relaxation	GSR (1)	1	2.78%
	Stress (Cognitive)	Fusion (ECG, OCU), GSR(1)	2	5.56%
	Stress (Physiological)	BS (1), Fusion(GSR, ECG) (1)	2	5.56%
Fatigue			2	5.56%

Mechanisms	Fatigue (Cognitive)	OCU (1)	1	2.78%
	Fatigue (Physical)	EMG (1)	1	2.78%
Error Processing Mechanisms			2	5.56%
	Error Potentials	EEG (2)	2	5.56%
Unclear/ Inferred			3	8.33%
	Unclear/Inferred	EEG (2)	2	5.56%
Total			36	100%

Error Processing Mechanisms were targeted in 2 manuscripts (5.56%). Both papers (Kim et al., 2020; Schiatti et al., 2018) used EEG to measure error related potentials. The neurophysiological responses to unusual or wrong observable actions of robots were used in the reward function of the robot during the learning process.

Finally, the targeted mechanisms were unclear for 2 manuscripts (5.56%). For example, (Labonte-Lemoyne et al., 2018), with EEG, measured the parietal upper alpha band to adapt Tetris game speed. However, the psychophysiological inferences are not clearly defined.

Neuro-adaptive systems leverage the targeted neurophysiological mechanisms and the psychophysiological inferences targeted to drive adaptation. Table 14 presents how the manuscripts utilize the mechanisms to implement different types of controls during the task. We found various unisensory (e.g., visual, auditory, haptic) and multisensory (e.g., visual, and auditory) feedback. 18 manuscripts (50%) built artifacts that provided sensory feedback of some form. 9 of them focused on attentional mechanisms for developing the neuro-adaptive system. For example, in a dashboard task, Toreini, Langner et Maedche (2020) used dwell time on the different metrics to infer the visual allocation of attention. The neuro-adaptive system directly provided to which metrics the user concentrated its gaze. The authors observed that this visual feedback enables users to manage their focus more efficiently. Demazure et al. (2019) concentrated on attentional mechanisms using EEG in a dashboard monitoring task. The manuscripts evaluated the efficacy of visual feedback in supporting sustained attention during a long-duration logistic task and measured the effect on behaviors and performance. Other mechanisms were used to drive sensory feedback. For example, building on information processing mechanisms, Breslow et al. (2014) modeled users' overload based on gaze behavior to drive visual cues in an unmanned vehicles supervision task. Those cues were designed to redirect users' attention toward potential threats and to support the user during overload segments to reduce the risk of errors. This type of adaptation could qualify as biofeedback; they redirect information about their states to the user.

Table 14

Mechanisms and neuro-adaptive System Control

Neuro-adaptive systems control	Mechanism Targeted								
	Att. M.	Aff. M	S. M.	I.P. M	E.P. M	F. M.	Unclear Inferred	Count	%
Visual feedback	6	2	3	1		1		13	36.11%
Auditory feedback	1							1	2.78%
Haptic feedback					1			1	2.78%
Visual/Auditory feedback	1		1					2	5.56%
Visual/Auditory/Haptic feedback	1							1	2.78%
Task load adaptation	1	2	1	3			2	9	25.00%
Task allocation	2	1				1		4	11.11%
System behavior change	1	1			1			3	8.33%
Neuromodulation	1							1	2.78%
Unclear/inferred				1				1	2.78%
Total	14	6	5	5	2	2	2	36	100%

Note. Attentional Mechanisms (Att. M.), Affective Mechanisms (Aff. M), Information Processing Mechanisms (I.P. M.), Stress Mechanisms (S. M.), Fatigue Mechanisms (F. M.), Error Processing Mechanisms (E.P. M.), Fatigue Mechanisms (F. M.)

However, neuro-adaptive systems were sometimes designed to control the characteristics of the HCI tasks. We noted three types of systems control: task load adaptation (9, 25%), task allocation (4, 11.11%), and system behavior change (3, 8.33%). In the case of task adaptation, the neuro-adaptive systems had control over specific parameters of the tasks but did not change their nature. For example, Labonte-Lemoyne et al. (2018) adjusted the speed of a game using alpha power bands compared to a baseline, i.e., a rest state. Zhou et al. (2015) inferred cognitive load with GSR to adapt the task workload dynamically to its users. In the context of air traffic control, Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al. (2016) built an artifact that adapts the load level of an air traffic monitoring system based on the mental workload of the users measured via EEG. The adaptation rules are built to adjust task cognitive demand to support the user by lowering the mental workload. For example, in case of high workload, the system would activate critical alerts in the interface, or the system will preselect and display aircraft to display on the system.

Another observed neuro-adaptive system control over tasks is task allocation, where the system controls the nature of the task. Some studies have utilized psychophysiological inferences to adjust the level of automation. In these cases, the neuro-adaptive system triggers either partial or complete task takeover between the automation and the user. For example, Di Flumeri et al. (2019) built an artifact to adjust the level of automation between two levels based on the vigilance of air traffic controllers measured by EEG. The neuro-adaptive system distributes the task load between the operator and the machine. The artifact operates under the assumption that automation reduces vigilance and switching to manual control maintains engagement. In a pilot test, Ramos et al. (2021) modeled the emotional states of a drone operator and created its digital twin. The predicted state of the twin would serve to gate commands to the drone and order it to stabilize autonomously. A neuro-adaptive system for physical task allocation between a robotic hand and an operator was developed in another study (Peternel et al., 2018). During the task, the robotic arms learn the operator's behavior to replicate it. Once the fatigue level is deemed too high, the neuro-adaptive system triggers the machine's takeover of the task.

We have classified 3 (8.33%) manuscripts under the classification of "system behavior change". This type of control is defined as an adaptation triggered by a neuro-adaptive system that continuously modifies the behavior of the system. This dynamic adaptation can be either covert or overt to the users. As examples of covert behavior adaptation, Govindarajan et al. (2018) integrated affective state estimation of the driver into a driver assistance system to predict breaking reaction time. The driver assistance system triggers the collision alerts based on the predicted reaction time, vehicle, and environmental parameters. Causse et al. (2019) created a decision support system for route suggestions in air traffic control. The systems adaptively adjust the decision system's parameters based on the users' fixations during the surveillance task. In both cases, the neuro-adaptive systems covertly modify the system's behavior for its user. On the other hand, one manuscript presented an artifact that overtly adapts a robot's behavior based on EEG. Kim et al. (2020) used error-related potential detection in a reinforcement learning protocol for robotic hand action execution.

2.5 Discussion

Our scoping review provides an overview of research on neuro-adaptive systems applied in HCI. We identified relevant primary literature by searching seven databases in diverse domains. However, given the emerging nature of this research, we did not assess the methodological quality of individual studies. We coded the manuscripts based on a framework adapted from design science (Gregor & Hevner, 2013; Hevner et al., 2004; Venable, 2006) and the neuro-adaptive system literature (Mason & Birch, 2003; Van Gerven et al., 2009). We mapped the neuro-adaptive artifacts using the three spaces of problem, knowledge, and solution. Based on this descriptive review, we report the current challenges and gaps in the literature.

2.5.1 Overview of the neuroadaptive system literature in HCI

Our results confirm the emergent nature of the research field. Over the last decade, we have observed a significant increase in journal and conference manuscripts related to this topic, indicating a growing momentum in this study area. The included manuscripts in our review appear to vary in their approach to defining the problem and motivating the artifact's creation. Only three studies (8.33%) clearly stated their research questions, while the artifact objective was clearly defined in 22 papers (61.11%). A qualitative analysis of the motivation behind the research showed that the need for a neuro-adaptive system emerged from various factors such as the task (e.g., automation, task demand), the user (e.g., stress), or the neuro-adaptive system (e.g., biofeedback design). Despite this variation, designers commonly linked these factors to their impact on user states, such as workload, motivation, and fatigue. Additionally, the consequences of how these factors impact users were mainly centered around the user's learning, health, and experience, or the task's performance. These results offer an interesting perspective on how neuro-adaptive systems are motivated in the current literature, highlighting two distinct approaches in which considering user states during human-computer interaction can either benefit users personally or aid in accomplishing tasks.

The literature shows diverse designs of neuro-adaptive systems and their components. When only one neurophysiological measure is employed, the most used measures of the users and inputs of the artifact were EEG (38.89%), followed by GSR (11.11%) and oculometry (11.11%). However, a significant proportion of the prototypes used sensor fusion techniques to leverage multiple measurement instruments (22.22%). Then, features in both the time and frequency domains were computed for two primary uses: their aggregation into indexes or their compilations into feature vectors. The use of indexes was diverse. Indexes translated the neurological signal into logical inputs for the systems using thresholds, fuzzy computational methods, and inputs for machine learning models. On the other hand, feature vectors only serve machine learning techniques to classify the state of the users. SVM and Linear Discriminant Analysis were popular choices for learning algorithms.

Psychophysiological inferences and neurophysiological mechanisms are employed to drive controls of neuro-adaptive systems. We observe two categories of control approaches in the literature: sensory feedback provided to users and adaptations made to various aspects of HCI tasks. In the first category, the user interface incorporates feedback modalities like visual, auditory, and haptic cues, often capitalizing on attention mechanisms. For instance, visual feedback can be used to redirect users' attention towards potential threats. The second category encompasses a wider range of targeted neurophysiological mechanisms, but the underlying goal remains system control in HCI tasks, including task load adaptation, task allocation, and system behavior modification. As an example, these systems can adjust the level of automation between the operator and the machine according to the user's vigilance and fatigue levels. Furthermore, they can continuously adapt the system's behavior, such as adjusting decision system parameters based on users' fixations during surveillance tasks.

The results reveal a growing body of literature on applied neuro-adaptive systems, which employ diverse neurophysiological measures, measurement techniques, and control designs. However, this emerging field also faces several challenges that need to be addressed. This section will discuss these challenges and propose potential solutions.

2.5.2 Challenges and guidelines of the literature

2.5.2.1 Defining the problem space

In manuscripts that aim to design and evaluate neuro-adaptive systems within HCI, it is essential to define the problem clearly. The necessity for a neuro-adaptive system arises from a specific problem that researchers or designers seek to address. Defining this problem is crucial in establishing the artifact's relevance and purpose (Gregor & Hevner, 2013). However, we observed that the included manuscripts often lack clarity on this aspect.

The current scarcity of guiding design frameworks and principles specifically tailored for passive neuro-adaptive systems is a potential reason for this issue. Fairclough and Lotte (2020) highlighted that this area remains less explored than the more established field of active control BCI design. Researchers have attempted to address this challenge. For instance, Mason and Birch (2003) proposed a functional model for BCI design. However, their framework primarily focuses on the components of the artifacts and does not directly address the problem definition. The authors made a commendable effort to define the artifact, yet their approach seems to encompass only specific aspects of the problem definition. The framework covers the users, the tasks, and the environments (e.g., target population, target tasks) with components of the neuro-adaptive systems, which muddles the differences between the problem definition and the designed artifact that solves it.

We posit that these aspects must be addressed within the problem space to clearly define the purpose, scope, nature, and relevance of the problem. For example, in this scoping review, we conceptualized the problem space as the environment from which the problem arises (i.e., the industry/domain, the target users, and the target task) and the problematization (i.e., the problem's relevance, research goal/question, and the neuro-adaptive artifact's objective). In this case, "target users" and "target tasks" refer to the prospective users and tasks of the neuro-adaptive system, which may not initially be the ones with which the artifact is evaluated. Nonetheless, these components are essential for effectively communicating the contribution of the proposed design.

In summary, manuscripts should clearly articulate the problem to be addressed with neuro-adaptive artifacts when motivating the research. While our conceptualization of the problem space can be used as a reference, researchers are encouraged to develop design frameworks for neuro-adaptive systems that specifically provide prescriptive requirements to address this challenge effectively.

2.5.2.2 Anchoring psychophysiological inference and constructs in the understanding of the body and the brain

Psychophysiological inferences constitute a significant challenge for neuro-adaptive artifacts. Neuro-adaptive systems inherently build on psychophysiological inferences (Fairclough, 2009) as the artifact relies on recognizing patterns in the sensor signal caused by mental states (Van Gerven et al., 2009). However, we found that the current literature lacks in anchoring the inferences in the current understanding of the body and the brain. This can lead to consequences that might hinder the designed artifacts from attaining their objectives. We found undefined psychophysiological constructs and various similar constructs related to similar neurophysiological mechanisms. It can also lead to misalignment between the psychophysiological inferences, targeted patterns, and the task. Finally, it can impede the ability of future neuro-adaptive artifacts to build on existing inferences.

To illustrate an example of multiple constructs that refer to the same mechanisms, we identified four constructs associated with information processing mechanisms: cognitive load (Yuksel et al., 2016), cognitive workload (Zhou et al., 2015), and mental workload (Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016; Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016) and overload (Breslow et al., 2014). These constructs draw upon different psychological theories, such as cognitive load theory and working memory (Baddeley, 1992; Sweller, 1994). Nevertheless, all of these neuro-adaptive artifacts are based on similar assumptions: users possess limited cognitive and attentional capacities, and performance declines when task demands exceed these capacities. The targeted mechanism is related to information processing processes and working memory (Baddeley, 1992). Notably, three of these manuscripts relied on neurophysiological measurements using EEG and fNIRS, while the others employed GSR and oculometry, demonstrating that researchers utilize instruments other than brain measurements. We found a similar situation with the attentional mechanisms and psychophysiological inference, a challenge already discussed extensively in the literature (Oken et al., 2006). This example highlights the diversity of explanations for closely related constructs, even when the underlying mechanisms are strongly interconnected.

In another case, there was a misalignment between the psychophysiological inferences and the task. For example, Govindarajan et al. (2018) relied on EEG and TFI to characterize a car driver's "affective state" to adapt the parameters of the car's assistance system. However, they labeled training data on cognitive workload level using an n-back task. This paradigm was created to manipulate working memory (Kirchner, 1958). Another manuscript avoided making psychophysiological inferences, mainly focusing on the neurophysiological patterns targeted (i.e., alpha activity in EEG) to adapt the speed of a game without clearly relating it to a clear and defined construct (Labonte-Lemoyne et al., 2018).

The literature review highlights the challenges of building upon psychophysiological inferences to create neuro-adaptive systems. Making psychophysiological inferences is difficult (Fairclough, 2009) as it requires establishing that a specific task elicits a particular change in the signal, even though the manipulation may induce multiple concurrent changes in the body and brain, which can vary across users (Fairclough, 2009). This challenge is even more significant in naturalistic tasks, such as those in HCI, as the task itself can induce simultaneous changes in users' mental states. Nonetheless, anchoring physiological inferences in our understanding of the body and brain can enhance the efficiency of neuro-adaptive artifacts. For example, understanding the brain can improve the creation of features from the signal and increase the classification models' performance (Van Gerven et al., 2009). Nicolelis and Lebedev (2009) even argue that neuro-adaptive systems can enhance our ability to test hypotheses and understand neurophysiological phenomena. Of course, it is essential to acknowledge that most manuscripts use correlates of the targeted users' mental states. However, manuscripts should still clearly define constructs and the psychophysiological inferences (i.e., the relationship between the physiological patterns and the psychological constructs) on which the neuro-adaptive system is built before describing the features extracted. The artifact's description should describe the neurophysiological mechanisms used.

2.5.2.3 Building on the state-of-the-art in decoding users' states

The scoping review highlighted that the current literature does not make use of state-of-the-art machine learning algorithms for neurophysiological signal classification. For instance, several papers employ feature-based approaches using popular models such as SVM, variations of LDA, or Random Forest, which are widely accepted (Lotte et al., 2018). However, there have been significant advancements in the use of adaptive classifiers (e.g., adaptive LDA, QDA, dynamic SVM, adaptive Gaussian classifiers), shrinkage LDA, or Riemannian Minimum Distance to the Mean, which are known to perform well in the context of limited data (Lotte et al., 2018), but seems to be underutilized. Deep learning techniques are also scarcely used despite their potential to provide better generalization and flexibility for neurophysiological data (Roy et al., 2019). Although there are still performance concerns with deep learning techniques, compared to the above techniques, Lotte et al. (2018) anticipate that their use may be relevant for end-to-end domain adaptation and data augmentation techniques, such as generative adversarial networks. Both techniques are highly pertinent for neuro-adaptive systems as the research aims to enhance task generalizability in HCI, which is frequently built on small datasets.

Although we acknowledge the importance of evaluating state-of-the-art mental state classification techniques for neuro-adaptive systems development, it is also crucial to recognize that linear classifiers have demonstrated satisfactory performance (Van Gerven et al., 2009). The primary area for improvement lies in the feature design and selection process. For instance, some papers in the review have presented modified versions of linear discriminant analysis incorporating automatic feature selection techniques (Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016; Gianluca Di et al., 2019). However, this approach, which is focused on the data, performs well when trained and tested on the same users but may be limited in its ability to generalize to multiple users due to inherent differences in physiological responses across individuals. In such cases, deep learning techniques, particularly end-to-end approaches, may offer a solution to the issue of generalization performance (Roy et al., 2019).

2.5.2.4 Acknowledging ethical considerations

The ethical implications of transferring neuro-adaptive systems from the laboratory to consumer markets are vital. Our review reveals that the literature focuses on workplace and critical job applications. For example, several papers have investigated the use of neuro-adaptive systems in air traffic control (Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016; Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016; Causse et al., 2019; Di Flumeri et al., 2019) or business domains (Demazure et al., 2019; Toreini et al., 2020). However, none of the papers included in this review explicitly address ethical concerns despite their aim of designing neuro-adaptive systems for use in work environments or consumer applications.

It is imperative for designers to consider future barriers and anticipate ethical issues that may impede the widespread adoption of neuro-adaptive systems. These systems pose significant challenges in this regard. (Fairclough, 2009) raises concerns regarding privacy issues and questions about who owns the physiological data such systems collect. Additionally, should the user be informed if the signal indicates the presence of a potential disease (e.g., cardiovascular risks, hypertension)? Furthermore, the author extends this notion to the public use of these technologies. Neuro-adaptive systems utilized in the workplace could inadvertently reveal personal information, such as covert emotional responses to colleagues around the user. These systems also raise questions regarding user autonomy or self-determination. In this regard, Reynolds and Picard (2005, p. 2) directly address systems designers by asking, "Could a user be emotionally manipulated by a program with the capability to recognize and convey affect? […] Should an affective system try to change the emotional state of a user?" Designers and researchers must address these ethical concerns to ensure the responsible development and implementation of neuro-adaptive systems.

In conclusion, our review of applied research on neuro-adaptive systems in HCI has revealed that ethical challenges are not being adequately addressed. Although discussions of these issues exist (Fairclough, 2009; Reynolds & Picard, 2005; Van Gerven et al., 2009), they need to be translated into practical implications for designers and researchers of such systems. We argue that there is a gap in the current literature regarding the actual requirements for neuro-adaptive systems design that would consider ethical and moral concerns specific to the artifact. To address this gap, an ethical space could be added to our current analytical framework. This space would refer to the characteristics of the neuro-adaptive artifact that could give rise to moral and ethical concerns for the designer, user, or organization.

2.5.3 Challenges and guidelines of the literature

This scoping review employed rigorous and transparent methods throughout the entire process. Firstly, the review protocol was developed following the methodology outlined by Arksey and O'Malley (2005) and followed the PRISMA extension for scoping reviews checklist (Tricco et al., 2018). To ensure a comprehensive literature search, we conducted searches across seven databases, carefully selecting multidisciplinary and specialized databases in medicine, business, psychology, and engineering. Secondly, for data charting, we adapted a central framework for design science research (Gregor & Hevner, 2013; Hevner et al., 2004; Venable, 2006), a field dedicated to constructing socio-technical artifacts (Gregor & Hevner, 2013), a relevant perspective for neuro-adaptive systems. Thirdly, since the design choices in neuro-adaptive systems are highly interdependent, we described the components of the artifacts in relation to other choices (i.e., sensors and features, features extractions and translators, targeted mechanisms, and systems control). This approach facilitated a coherent and fluid description of the artifacts throughout the results section. Fourthly, we identified and described three significant challenges in the design of neuro-adaptive systems for HCI and offered guidelines to inform future design and research opportunities in this domain.

This analysis lays the foundation for constructing a conceptual framework to advance the integration of neurophysiological measures into IS artifacts, commonly known as neuro-adaptive systems. By delineating the problem, knowledge, and solution spaces concerning neuro-adaptive systems and their components (such as functional elements and neuropsychological inferences), this review becomes a crucial step toward establishing a theoretical basis for designing these artifacts. Ultimately, this approach facilitates the accumulation of descriptive and prescriptive knowledge within the field of NeuroIS.

This review possesses some limitations. Firstly, although comprehensive, our search could have included more databases from the engineering field, which might result in a slight bias in the findings. While we included IEEExplore, we were unable to incorporate ACM due to the database's technical limitations. Secondly, we did not assess the quality of the manuscripts included in the scoping review, contrary to the recommendations of Daudt et al. (2013). Neuro-adaptive research in HCI still faces numerous challenges (Fairclough & Lotte, 2020). Our rationale for not conducting a quality assessment was that research outside active and clinical neuro-adaptive systems remains relatively nascent, and excluding research might conceal the domain's challenges. Thirdly, the initial search was conducted in March 2022; hence, an updated search should be performed to enhance the review's timeliness before publication.

2.6 Conclusion

In this manuscript, our objective is to understand the current state of the emerging field of neuro-adaptive system design in HCI, which could further inform mental state estimation in naturalistic IS tasks. This scoping review examines the existing literature on neuro-adaptive systems through a design science research framework. The findings illustrate a diverse body of literature that confronts various challenges. We discuss these challenges and offer guidance for researchers and designers in the field. In summary, we argue that the problems neuro-adaptive systems aim to address should be better conveyed. In terms of mental state estimations, researchers should describe the psychophysiological inferences upon which the artifact is constructed. State-of-the-art methods for decoding user states should be employed and evaluated (e.g., end-to-end deep learning for mental state decoding). Lastly, ethical considerations regarding the created artifact must be discussed to engage in with potential ethical issues the field may encounter. We hope that the results of our scoping review assist researchers and designers in developing neuro-adaptive systems and robust mental state inferences.

References

Appriou, A., Cichocki, A., & Lotte, F. (2018). Towards robust neuroadaptive HCI: exploring modern machine learning methods to estimate mental workload from EEG signals. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*,

Aranyi, G., Pecune, F., Charles, F., Pelachaud, C., & Cavazza, M. (2016). Affective Interaction with a Virtual Character Through an fNIRS Brain-Computer Interface. *Front Comput Neurosci*, *10*, 70. https://doi.org/10.3389/fncom.2016.00070

Arico, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J. P., Granger, G., Benhacene, R., & Babiloni, F. (2016). Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Front Hum Neurosci*, *10*, 539. https://doi.org/10.3389/fnhum.2016.00539

Arico, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., & Babiloni, F. (2016). A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. *Prog Brain Res*, *228*, 295-328. https://doi.org/10.1016/bs.pbr.2016.04.021

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International journal of social research methodology*, *8*(1), 19-32.

Azgomi, H. F., Cajigas, I., & Faghih, R. T. (2021). Closed-Loop Cognitive Stress Regulation Using Fuzzy Control in Wearable-Machine Interface Architectures. *Ieee Access*, *9*, 106202-106219. https://doi.org/10.1109/Access.2021.3099027

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556-559.

Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L., & Bekhuis, T. (2016). De-duplication of database search results for systematic reviews in EndNote. *Journal of the Medical Library Association: JMLA*, *104*(3), 240.

Breslow, L. A., Gartenberg, D., McCurry, J. M., & Trafton, J. G. (2014). Dynamic Operator Overload: A Model for Predicting Workload During Supervisory Control. *Ieee Transactions on Human-Machine Systems*, *44*(1), 30-40. https://doi.org/10.1109/Tsmc.2013.2293317

Causse, M., Lancelot, F., Maillant, J., Behren, J., Cousy, M., & Schneider, N. (2019). Encoding decisions and expertise in the operator's eyes: Using eye-tracking as input for system adaptation. *International Journal of Human-Computer Studies*, *125*, 55-65. https://doi.org/10.1016/j.ijhcs.2018.12.010

Chaouachi, M., Jraidi, I., & Frasson, C. (2015). Adapting to learners' mental states using a physiological computing approach. *Proceedings of the 28th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2015*, 257-262.

Darzi, A., & Novak, D. (2021). Automated affect classification and task difficulty adaptation in a competitive scenario based on physiological linkage: An exploratory study. *Int J Hum Comput Stud*, *153*, N.PAG-N.PAG. https://doi.org/10.1016/j.ijhcs.2021.102673

Daudt, H. M., van Mossel, C., & Scott, S. J. (2013). Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC medical research methodology*, *13*(1), 1-9.

Demazure, T., Karran, A., Labonte-LeMoyne, E., Leger, P. M., Senecal, S., Fredette, M., & Babin, G. (2019). Sustained Attention in a Monitoring Task: Towards a Neuroadaptative Enterprise System Interface. *Information Systems and Neuroscience (Neurois Retreat 2018)*, *29*, 125-132. https://doi.org/10.1007/978-3-030-01087-4_15

Demazure, T., Karran, A., Léger, P.-M., Labonté-LeMoyne, É., Sénécal, S., Fredette, M., & Babin, G. (2021). Enhancing Sustained Attention. *Business & Information Systems Engineering*, *63*(6), 653-668. https://doi.org/10.1007/s12599-021-00701-3

Demazure, T., Karran, A. J., Labonté-LeMoyne, É., Léger, P.-M., Sénécal, S., Fredette, M., & Babin, G. (2018). Sustained attention in a monitoring task: Towards a neuroadaptative enterprise system interface. *Information Systems and Neuroscience*, Chapter 15.

Dey, A., Chatburn, A., & Billinghurst, M. (2019). Exploration of an EEG-Based Cognitively Adaptive Training System in Virtual Reality. *2019 26th Ieee Conference on Virtual Reality and 3d User Interfaces (Vr)*, 220-226. https://doi.org/10.1109/VR.2019.8797840

Di Flumeri, G., De Crescenzio, F., Berberian, B., Ohneiser, O., Kramer, J., Arico, P., Borghini, G., Babiloni, F., Bagassi, S., & Piastra, S. (2019). Brain-Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems. *Front Hum Neurosci*, *13*, 296. https://doi.org/10.3389/fnhum.2019.00296

El-Samahy, E., Mahfouf, M., Torres-Salomao, L. A., & Anzurez-Marin, J. (2015). A New Computer Control System for Mental Stress Management using Fuzzy Logic. *2015 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, 1-7. https://doi.org/10.1109/EAIS.2015.7368785

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

Fairclough, S. H., & Lotte, F. (2020). Grand Challenges in Neurotechnology and System Neuroergonomics [Specialty Grand Challenge]. *Frontiers in Neuroergonomics*, *1*. https://doi.org/10.3389/fnrgo.2020.602504

Ghandi, M., Blaisdell, M., & Ismail, M. (2021). Embodied empathy: Using affective computing to incarnate human emotion and cognition in architecture. *International Journal Of Architectural Computing*, *19*(4), 532-552.

Gianluca Di, F., De Crescenzio, F., Berberian, B., Ohneiser, O., Kramer, J., Aricò, P., Borghini, G., Babiloni, F., Bagassi, S., & Piastra, S. (2019). Brain–Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems. *Frontiers in Human Neuroscience*. https://doi.org/http://dx.doi.org/10.3389/fnhum.2019.00296

Gorjan, D., Gramann, K., De Pauw, K., & Marusic, U. (2022). Removal of movement-induced EEG artifacts: current state of the art and guidelines. *Journal of neural engineering*.

Govindarajan, V., Driggs-Campbell, K., & Bajcsy, R. (2018). Affective Driver State Monitoring for Personalized, Adaptive ADAS. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 1017-1022. https://doi.org/10.1109/ITSC.2018.8569585

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, *37*(2).

Gu, X., Cao, Z., Jolfaei, A., Xu, P., Wu, D., Jung, T.-P., & Lin, C.-T. (2021). EEG-based brain-computer interfaces (BCIs): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications. *IEEE/ACM transactions on computational biology and bioinformatics*, *18*(5), 1645-1666.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, *28*(1), 75-105.

Karthikeyan, R., & Mehta, R. K. (2020). Towards a Closed-Loop Neurostimulation Platform for Augmenting Operator Vigilance. *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3976-3983. https://doi.org/10.1109/SMC42975.2020.9283126

Kim, S. K., Kirchner, E. A., & Kirchner, F. (2020). Flexible online adaptation of learning strategy using EEG-based reinforcement signals in real-world robotic applications. *2020 Ieee International Conference on Robotics and Automation (Icra)*, 4885-4891. https://doi.org/10.1109/ICRA40945.2020.9197538

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*, *55*(4), 352.

Kline, J. E., Huang, H. J., Snyder, K. L., & Ferris, D. P. (2015). Isolating gait-related movement artifacts in electroencephalography during human walking. *J Neural Eng*, *12*(4), 046022. https://doi.org/10.1088/1741-2560/12/4/046022

Labonte-Lemoyne, E., Courtemanche, F., Louis, V., Fredette, M., Sénécal, S., & Léger, P.-M. (2018). Dynamic threshold selection for a biocybernetic loop in an adaptive video game context. *Frontiers in Human Neuroscience*, *12*, 282.

Larradet, F., Barresi, G., & Mattos, L. S. (2017). Effects of galvanic skin response feedback on user experience in gaze-controlled gaming: A pilot study. *Annu Int Conf IEEE Eng Med Biol Soc*, *2017*, 2458-2461.

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: advancing the methodology. *Implementation Science*, *5*(1), 1-9.

Lim, Y., Pongsakornsathien, N., Gardi, A., Sabatini, R., Kistan, T., Ezer, N., & Bursch, D. J. (2021). Adaptive Human-Robot Interactions for Multiple Unmanned Aerial Vehicles. *ROBOTICS*, *10*(1), 1-30.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, *15*(3), 031005.

Lotte, F., & Roy, R. N. (2019). Brain–Computer Interface Contributions to Neuroergonomics. In *Neuroergonomics* (pp. 43-48). Elsevier.

Lux, E., Adam, M. T., Dorner, V., Helming, S., Knierim, M. T., & Weinhardt, C. (2018). Live biofeedback as a user interface design element: A review of the literature. *Communications of the Association for Information Systems*, *43*(1), 18.

Mason, S. G., & Birch, G. E. (2003). A general framework for brain-computer interface design. *IEEE Trans Neural Syst Rehabil Eng*, *11*(1), 70-85. https://doi.org/10.1109/TNSRE.2003.810426

Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019). Are we ready for real-world neuroscience? , *31*(3), 327-338.

Nalepa, G. J., Kutt, K., Gizycka, B., Jemiolo, P., & Bobek, S. (2019). Analysis and Use of the Emotional Context with Wearable Devices for Games and Intelligent Assistants. *Sensors (Basel)*, *19*(11), 2509. https://doi.org/10.3390/s19112509

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *Neuroimage*, *222*, 117254.

Nicolelis, M. A., & Lebedev, M. A. (2009). Principles of neural ensemble physiology underlying the operation of brain–machine interfaces. *Nature reviews neuroscience*, *10*(7), 530-540.

Oken, B. S., Salinsky, M. C., & Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin Neurophysiol*, *117*(9), 1885-1901. https://doi.org/10.1016/j.clinph.2006.01.017

Parnandi, A., & Gutierrez-Osuna, R. (2021). Partial Reinforcement in Game Biofeedback for Relaxation Training. *IEEE Transactions on Affective Computing*, *12*(1), 141-153. https://doi.org/10.1109/Taffc.2018.2842727

Pavlidis, I., Khatri, A., Buddharaju, P., Manser, M., Wunderlich, R., Akleman, E., & Tsiamyrtzis, P. (2021). Biofeedback Arrests Sympathetic and Behavioral Effects in Distracted Driving. *IEEE Transactions on Affective Computing*, *12*(2), 453-465. https://doi.org/10.1109/Taffc.2018.2883950

Peternel, L., Tsagarakis, N., Caldwell, D., & Ajoudani, A. (2018). Robot adaptation to human physical fatigue in human-robot co-manipulation. *AUTONOMOUS ROBOTS*, *42*(5), 1011-1021. https://doi.org/10.1007/s10514-017-9678-1

Raaijmakers, S. F., Steel, F. W., de Goede, M., van Wouwe, N. C., van Erp, J. B. F., & Brouwer, A. M. (2013). Heart rate variability and skin conductance biofeedback: A triple-blind randomized controlled study. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (Acii)*, 289-293. https://doi.org/10.1109/Acii.2013.54

Ramos, D., Goncalves, G., Faria, R., & Sanches, M. P. (2021). Building a Drone Operator Digital Twin using a Brain-Computer Interface for Emotion Recognition. *2021 20th International Conference on Advanced Robotics (ICAR)*, 824-829. https://doi.org/10.1109/Icars3236.2021.9659360

Reynolds, C., & Picard, R. W. (2005). Evaluation of affective computing systems from a dimensional metaethical position. First Augmented Cognition International Conference, Las Vegas, NV,

Riedl, R., & Léger, P.-M. (2016). Fundamentals of NeuroIS. *Studies in Neuroscience, Psychology and Behavioral Economics. Springer, Berlin, Heidelberg*.

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, *16*(5), Article 051001. https://doi.org/10.1088/1741-2552/ab260c

Schiatti, L., Tessadori, J., Deshpande, N., Barresi, G., King, L. C., & Mattos, L. S. (2018). Human in the Loop of Robot Learning: EEG-based Reward Signal for Target Identification and Reaching Task. *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 4473-4480. https://doi.org/10.1109/ICRA.2018.8460551

Stangl, M., Maoz, S. L., & Suthana, N. (2023). Mobile cognition: imaging the human brain in the 'real world'. *Nature reviews neuroscience*, 1-16.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295-312.

Szafir, D., & Mutlu, B. (2012). Pay attention! Designing adaptive agents that monitor and improve user engagement. *Conference on Human Factors in Computing Systems - Proceedings*, 11-20. https://doi.org/10.1145/2207676.2207679

Toreini, P., Langner, M., & Maedche, A. (2020). Using Eye-Tracking for Visual Attention Feedback. *Information Systems and Neuroscience*, *32*, 261-270. https://doi.org/10.1007/978-3-030-28144-1_29

Toreini, P., Langner, M., Maedche, A., Morana, S., & Vogel, T. (2022). Designing Attentive Information Dashboards. *Journal of the Association for Information Systems*, *2021*.

Trachel, R. E., Brochier, T. G., & Clerc, M. (2018). Brain-computer interaction for online enhancement of visuospatial attention performance. *J Neural Eng*, *15*(4), 046017. https://doi.org/10.1088/1741-2552/aabf16

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D., Horsley, T., & Weeks, L. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*, *169*(7), 467-473.

Tseng, K. C., Wang, Y., Lin, B., & Hsieh, P. H. (2012). Brain Computer Interface-based Multimedia Controller. *2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 277-280. https://doi.org/10.1109/IIH-MSP.2012.73

Van Gerven, M., Farquhar, J., Schaefer, R., Vlek, R., Geuze, J., Nijholt, A., Ramsey, N., Haselager, P., Vuurpijl, L., & Gielen, S. (2009). The brain–computer interface cycle. *Journal of neural engineering*, *6*(4), 041001.

Venable, J. (2006). The role of theory and theorising in design science research. *Proceedings of the 1st International Conference on Design Science in Information Systems and Technology* (DESRIST 2006),

Veritas Health Innovation. (2023). *Covidence systematic review software*. In Veritas Health Innovation. www.covidence.org.

vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a neurois research agenda with four areas of societal contributions. *European Journal of Information Systems*, *29*(1), 9-24.

Vortmann, L. M., & Putze, F. (2020). Attention-Aware Brain Computer Interface to avoid Distractions in Augmented Reality. *Chi'20: Extended Abstracts of the 2020 Chi Conference on Human Factors in Computing Systems*.

Wang, Y., Shen, X. T., Liu, H. W., Zhou, T. T., Merilampi, S., & Zou, L. (2019). The Effectiveness of EEG-Feedback on Attention in 3D Virtual Environment.

Intelligent Robotics and Applications, Icira 2019, Pt V, 11744, 99-107. https://doi.org/10.1007/978-3-030-27541-9_9

Whelan, E., McDuff, D., Gleasure, R., & Vom Brocke, J. (2018). How emotion-sensing technology can reshape the workplace. *MIT Sloan Management Review*, *59*(3), 7-10.

Yuksel, B. F., Oleson, K. B., Harrison, L., Peck, E. M., Afergan, D., Chang, R., & Jacob, R. J. K. (2016). Learn Piano with BACh: An Adaptive Learning Interface that Adjusts Task Difficulty based on Brain State. *34th Annual Chi Conference on Human Factors in Computing Systems, Chi 2016*, 5372-5384. https://doi.org/10.1145/2858036.2858388

Zander, T. O., & Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of neural engineering*, *8*(2), 025005.

Zander, T. O., Kothe, C., Jatzev, S., & Gaertner, M. (2010). Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In *Brain-Computer Interfaces* (pp. 181-199). Springer.

Zargari Marandi, R., Madeleine, P., Omland, Ø., Vuillerme, N., & Samani, A. (2019). An oculometrics-based biofeedback system to impede fatigue development during computer work: A proof-of-concept study. *Plos One*, *14*(5), 1-24.

Zhang, Y., Xiao, H., Zhang, J., & Cai, H. (2021). An Adaptive Attention Regulation Method Based on Biocybernetic Loop. *2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM)*, 1-6. https://doi.org/10.1109/HEALTHCOM49281.2021.9399050

Zhou, J., Jung, J. Y., & Chen, F. (2015). Dynamic workload adjustments in human-machine systems based on GSR features. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9296*, 550-558. https://doi.org/10.1007/978-3-319-22701-6_40

Chapter 3
Essay #2 - End-to-end Deep Learning Approaches to Mental Workload Classification using Electroencephalography in HCI

Abstract

We present research towards a complete end-to-end deep learning process for mental workload estimation during natural Human-Computer Interaction (HCI) tasks. We utilized two sets of tasks, an n-back task to elicit different levels of mental workload (MW) and an ecologically valid HCI task developed to provide similar difficulty levels using a simulated environment. A review of mental workload estimation using deep learning approaches and EEG signal data revealed that fully utilizing end-to-end processes (i.e., no manual engineering of features) is still rare in the scientific literature. Hence, we aim to answer the following question: Can end-to-end deep learning be used to classify MW in a natural HCI task? We benchmarked several deep learning models and selected the two best-performing architectures (i.e., FCN, a fully convolutional neural network, and ResNet, a residual network). Based on the machine learning (ML) and neuroscience literature, architectural and parameter design choices specific to EEG signal data were implemented, and we present a systematic assessment of those. The two selected classifier models achieved an average accuracy of .933 (± .054) and .917 (± .074) for FCN and ResNet, respectively. We validate our models through rigorous assessments of their neurophysiological plausibility, robustness, and reliability. We then deployed these two classifier models on previously unobserved EEG data, intending to measure empirical evidence pertaining to the relationship between complexity, mental workload, and performance. Our classification findings revealed an inverse U-shaped relationship between complexity and estimated mental workload. This research further enriches the academic discourse by thoroughly appraising architectural and design considerations relevant to EEG-based mental workload classification in natural HCI settings.

3.1 Introduction

When exerting considerable mental effort during human-computer interaction (HCI), human information processing capabilities decrease in accordance with a mental workload (MW) increase, such that the ability to process and attend to incoming sensory and data streams attenuates, which may act to decrease situational awareness and impede effective decision-making abilities (Wickens, 2002). Thus, an individual's workload capacity can constrain their ability during cognitively demanding tasks. Potential solutions relate to improvements in interface design and training or to dynamically and continuously adapt interfaces to maintain mental workload at optimal levels depending on the task. As a consequence, the ability to measure mental workload (Ayaz et al., 2012; Burke et al., 2004) has become an important concern for research and its application in HCI (e.g., game design, education applications, evaluating user experience, flight simulator training) (Appriou et al., 2018; Solovey et al., 2014).

While interacting with computer interfaces, an increase of MW in excess of an individual's capacity can lead to a state of cognitive overload that can reduce learning ability (Sweller et al., 2019; Sweller et al., 1998), and hinder one's capacity to allocate attention to between multiple tasks (Thomas & Wickens, 2001). Attention is a critical and scarce resource for multitasking in an environment composed of multiple informational displays (Wickens et al., 2003). Therefore, identifying and managing MW levels can positively benefit HCI users and designers.

The development of neurophysiological indices for the assessment of MW has been concurrent with advancements in the sensor technologies necessary to measure it. These measures can be utilized in various ways to evaluate the naturalness and efficacy of human-technology interfaces, monitor user responses to the computer-mediated task, or tailor the difficulty within the task. One popular technique for measuring MW is pupillometry. Variance in the diameter of the eye has been correlated with cognitive processing and activation of the brainstem (Laeng, Sirois, & Gredebäck, 2012). Numerous studies have used Pupillometry to measure mental workload in naturalistic contexts due to its non-intrusive nature (Buettner, 2013; Just et al., 2003; Klingner, Kumar, & Hanrahan, 2008; Palinko, Kun, Shyrokov, & Heeman, 2010). Other instruments, such as functional near-infrared spectroscopy (fNIRS) and electrocardiogram (ECG), have also been reported as effective sensors for measuring mental workload. Durantin et al. (2014) noted changes in oxygenation in the pre-frontal cortex when inducing cognitive overload (when controlling for difficulty and processing load) and in LF/HF (percentage ratio of low-frequency to high-frequency power) heart rate variability. It is possible to combine various sensor instruments and technologies to offer a more granular estimation of the mental workload. This process is often referred to as multi-modal or sensor fusion and provides a favourable trade-off regarding higher complexity in research design versus more robust MW measurement and assessment (Debie et al., 2019).

Chief among these technologies, electroencephalography (EEG) allows continuous monitoring of subjects' mental workload conveniently in the laboratory or the field (Lohani et al., 2019). Studies have found correlations between MW and variance in brainwaves expressed as increases or decreases in alpha (8-12hz) and theta (4-8hz) bands in pre-frontal brain regions (Grimes et al., 2008). The majority of approaches to mental workload estimation transform the signal from EEG into features within the time or frequency domain, such as extracting event-related potentials in the time domain (Blankertz et al., 2011); in the frequency domain using Fast Fourier Transform (FFT) (Jiao et al., 2018; Kwak et al., 2019); or hybrid Short-Time Fourier Transform (STFT) (Hefron, Borghetti, Schubert Kabban, et al., 2018; Kim et al., 2014) or in the time-frequency domain using wavelet transforms (WT) (Qayyum, Faye, et al., 2018; Qayyum, Khan, et al., 2018).

Unfortunately, signal transformation techniques and engineered features (especially images used as input) can be computationally expensive and impede online classification. Moreover, evidence shows that feature parameters (e.g., power band) might dynamically change with age, task demands, and cognitive states within and between individuals (Donoghue et al., 2020). For example, (Donoghue et al., 2021) show that methods using predefined frequency bands might fail to capture oscillatory activity between subjects accurately. This evidence led to the development of analytical methods that consider individual differences and highlight a significant limitation of engineered features (Donoghue et al., 2021).

End-to-end deep learning is a potential approach to address engineered feature limitations (Roy, Banville, et al., 2019). Deep learning algorithms are “representation-learning methods” that learn hierarchical representations of raw input data by decomposing the task into smaller non-linear problems (LeCun et al., 2015, p. 436). End-to-end models incorporate preprocessing feature extraction and discriminative classifiers in the learning phase. An end-to-end process in the context of deep learning for mental workload estimation describes a process that takes raw EEG signal data, processes these data, derives discriminant features, and then provides a classification of the target state as a complete functional solution. EEG processing and classification methodologies using deep learning techniques are experiencing rapid growth in research interest (Craik et al., 2019; Roy, Hubert, et al., 2019; Zhou et al., 2021, Chen et al., 2022).

Traditional approaches to classifying EEG data typically involve domain-specific processing methods, feature analysis, and supervised machine learning. These include Independent Component Analysis used for artefact removal (Jung et al., 2000), Principal Component Analysis (PCA), and local Fisher’s discriminant analysis applied for dimensionality reduction (Craik et al., 2019). For EEG feature classification and cognitive state estimation, supervised machine learning techniques often include Linear Discriminant Analysis (LDA), Riemannian Minimum Distance to Mean (RMDM), Support Vector Machines (SVM) in the researcher's arsenal of tools (Blankertz et al., 2011; Gao et al., 2019; Jiao et al., 2018; Lotte et al., 2018; Rojas et al., 2020).

Deep learning approaches offer novel yet complementary tools and have already shown promising results as part of an EEG processing pipeline (Craik et al., 2019). By design, deep learning techniques theoretically appear to provide a good fit for classifying the properties of EEG and addressing the processing challenges associated with EEG signals and deriving novel features that may potentially be more granular and discriminative than engineered features (Roy, Hubert, et al., 2019).

The research presented in this manuscript seeks to answer the question, "to what extent is it possible to estimate mental workload during naturalistic HCI tasks based on neurophysiological signal data using an end-to-end deep learning process?". The overarching objective is to create an offline deep learning model based on end-to-end processes for assessing and estimating mental workload during realistic HCI tasks in a high-fidelity simulator. In order to answer the research question, the objectives of this work are threefold: (i) Benchmark end-to-end deep-learning models for mental workload classification; (ii) Develop a mental workload classifier that achieves high classification performance; (iii) Estimate mental workload during a simulator task and replicate past empirical findings related to the relationship between task complexity, mental workload, and performance.

To begin, we reviewed the literature, revealing that complete end-to-end processes remain largely unused for EEG signal processing and classification of mental workload even though they provide some significant advantages. We demonstrate that specific end-to-end deep-learning architectures perform comparably to state-of-the-art baselines from the BCI literature. We show that a fully connected neural network and a residual neural network showed a significant increase in performance compared to a Riemannian geometry-based method, a technique that reaches state-of-the-art performance on recent BCI problems (Lotte et al., 2018). While these deep learning architectures presented superior performance, the lack of explainability related to the features used for training was a significant concern. Ensuring that a model does not learn from noise or systematic artefacts in the EEG signal is a concern raised in the current literature (Roy, Hubert, et al., 2019). To address this concern, we utilized an Integrated Gradients technique (Sundararajan, Taly et Yan, 2017) to validate the neurophysiological acceptability of the features learned by the models and then applied each model to new data and replicated previous findings regarding the relationship between mental workload and task complexity.

This manuscript makes a significant contribution to the estimation of mental workload, an important construct in HCI, as well as in the broader context of mental state estimations in HCI and NeuroIS. Estimating mental states during technology interaction holds the potential for valuable insights and contributions. Firstly, it can provide a deeper understanding of the cognitive mechanisms underlying the impact of technological artifacts on users' cognition and behavior (Riedl & Léger, 2016; vom Brocke et al., 2020). Consequently, mental state estimation can serve as a mediator between the artifact and user behavior (Riedl & Léger, 2016). Additionally, it can function as an evaluation tool in the design of artifacts (Riedl & Léger, 2016; vom Brocke et al., 2020). Lastly, real-time estimations can facilitate the development of neuro-adaptive systems (Riedl & Léger, 2016) further enhancing their design and functionality.

3.2 Literature Review

To better understand the state of the literature on the classification of mental workload with EEG using deep learning, we performed a literature review on the deep learning techniques used for mental workload decoding. To do so, we queried the "Web of Science" database using the following keywords: "deep learning," "neural network," with "working memory," "workload," "mental workload," "cognitive load," or "cognitive workload." To constrain our results, we conjoined terms to include the instrument of measurement: "electroencephalography" or "EEG".

The search resulted in 169 articles from conferences and journals, both manuscript types were included. We then filtered articles based on a publication year > 2010 to reduce the sample to more recent research, resulting in 124 "current" manuscripts. Manuscripts were screened to ensure the inclusion of the instruments ("EEG" only or "EEG" with other instruments), the phenomena of interest (the focal construct must be in the nomological network of "mental workload"). Additionally, the classification techniques ("deep learning" or "deep learning" with other approaches),based on authors, content and model similarity (if the dataset, methods, and results were similar in multiple papers). In the case of similarity, the journal or the conference article with the most information available was retained.

We extracted information from the 52 remaining articles related to model architectures, problem setting, task, preprocessing, tested design choices, and if the approach developed was end-to-end (data table is available in appendix). Table 15 shows the manuscripts leveraging end-to-end deep learning techniques to classify mental workload

Table 15

Manuscripts using fully end-to-end deep learning approaches for MW estimation with EEG

Authors	Problem Setting	Task (condition)	Inputs Domain	Data curation	Features	Model Name	Layer	Activation Function	Optimizers	Regularization	Training Strategies	Design choices	Baseline	Accuracy
Mohammad A. Almogbel et al., 2018	Driver workload level, with subject	Driving Simulator (dense traffic vs low traffic), within subject	Time, raw	No	**End-to-end**	CNN	7	Hidden layers: ReLu Output Layers: Softmax	RMSProp (lr =0.002)	Normalized by z-score	Overlapping windows at different time	Windows size # Layers	Deep learning with hand engineered features (1L CNN)	95.31%
Almogbel, Dang, Kameyama, et al., 2019	Driver workload level, within subject,	Driving Simulator (dense traffic vs low traffic vs zero traffic), within subject	Time, raw	No	**End-to-end**	CNN	8	Hidden layers: ReLu Output Layers: Softmax	RMSProp (lr=0.0001)	Normalized by z-score	Overlapping windows at different time	Windows size	Deep learning with hand engineered features (1L CNN)	96%
Gao et al., 2019	Driver fatigue, within-subject	Driving simulator. Highway	Time, raw	(1000 Hz, downsample 100 Hz), 1-50 Hz	**End-to-end**	EEG-based spatial–temporal convolutional neural network (ESTCNN)	14	Hidden: ReLu Output: Softmax	SGD (lr = 0.001)	Batch normalization	10 fold cross validation	Core block: three convolutional blocks and a pooling layer.	SVM, LSTM, CNNs, FFT + CNNs	97.37%
Hua et al., 2019	Working memory, within subject	Visuo-spatial working memory task	Time + Frequency, raw + values + images	frequency bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (15–30 Hz), low gamma (30–45 Hz) and all (0.5–45 Hz)	**End-to-end + FIR filters for rhythm extraction**	Brain connection based on CNN (BCCNN)						Type of inputs (fusions)		99%, 96.35%

3.2.1 Feature extraction in deep learning for ML estimation

During our review of the literature regarding deep learning approaches to mental workload estimation, we discovered that the majority of research approaches fail to leverage a major strength of the method, that is, the ability to learn discriminative features and produce classifications directly from the raw EEG signal. From 52 coded papers, only four papers used a fully end-to-end deep learning approach.

Our review revealed the widespread use of engineered features before being used as input to a deep learning classifier. Highly utilized transformations consisted of Fast Fourier Transform (Casson, 2014; Jiao et al., 2018; Qiao & Bi, 2020; Tao et al., 2019; Yang et al., 2019), Short-Time Fourier Transform (Hefron, Borghetti, Kabban, et al., 2018; Kim et al., 2014; Wang et al., 2012), or Wavelet Transform (Hefron et al., 2017; Qayyum, Khan, et al., 2018; Wu et al., 2019). Additionally, descriptive measures such as the mean, variance, skewness, and kurtosis of the power distribution, as well as the ratios between power bands, for example, Shannon entropy and spectral entropy (Hefron et al., 2017; Tao et al., 2019; Yang et al., 2019; Yin et al., 2019) also proved popular as input features for a model architecture.

Transformations such as these can be supplied to a neural network as values or transformed into graphical representations, thus changing the task into an image classification problem (Hefron, Borghetti, Schubert Kabban, et al., 2018; Jiao et al., 2018; Qayyum, Khan, et al., 2018; Qiao & Bi, 2020). Feature and domain-specific approaches were developed to solve problems inherent to EEG data (e.g., low signal-to-noise ratio, signal non-stationarity, inter-subject variability) and derive meaningful interpretations. However, the choices made when deriving features can have a considerable positive or negative impact on the performance of the classifiers and, further, any inferences made from classifications as a result.

Focusing our review efforts on the estimation of mental workload utilizing deep learning resulted in a low number of examples using end-to-end processes to discriminate between levels of mental workload directly from the sensor signals. However, the reported examples showed interesting results when utilizing end-to-end processes to train classifiers directly from the raw signal (Mohammad A Almogbel et al., 2018; Almogbel, Dang, & Kameyama, 2019). Further work showed promising results when estimating mental workload using models trained using end-to-end processes after applying minimum data correction, such as bandpass filtering (Gao et al., 2019; Hua et al., 2019).

End-to-end processes can offer a viable alternative to engineered features while reaching similar or improved performance to more traditional approaches. The main advantage of using end-to-end processes is that the resulting model is composed of hierarchical and translational invariant features that mitigate the problem of intra-class variability and inter-class similarities (Nweke et al., 2018). However, an issue that can arise when using EEG signals for mental workload estimation is when the levels of task difficulty and, thus, the neurophysiological responses are similar, such as when estimating moderate or high mental workload (Saadati et al., 2020b).

3.2.2 Architecture and design choices

Of the end-to-end approaches, the majority utilize a variation of convolutional networks (CNN). For example, Almogbel et al., 2018, 2019 leveraged CNNs with up to 8 layers. They made design choices such as ReLu as the activation function on the hidden layers and RMSProp as optimizers. Gao et al., 2019 developed an EEG-based spatial-temporal convolutional neural network (ESTCNN) of 14 layers, again using ReLu as the activation function and SGD as optimizers. Unfortunately, the reviewed papers show little consideration for explaining the design choices and their impacts.

Several techniques can be applied, and hyperparameters tuned to models before applying them to unseen data to improve the performance and generalization of deep learning models. In the following sections, we discuss design choices and how they impact performance with EEG based on machine learning literature and our review.

3.2.2.1 Optimization

Optimization techniques attempt to reduce the difference between training and test errors by minimizing the cost function of the model based on its parameters (Ruder, 2016). The current work aims to select an optimizer that provides fast convergence, stability and high performance over the EEG data and across participants. The selection of the optimizer is crucial for the construction of a deep learning model for EEG decoding. Unfortunately, such design choices have not yet been reported in the current literature (Roy, Banville, et al., 2019). In this manuscript, we systematically assess four optimizers: Adam (Kingma & Ba, 2014), SGD, Adadelta (Zeiler, 2012), and Nadam (Dozat, 2016), of which Adam and SGD are the two most used optimizers for deep learning applications on EEG data (Roy, Banville, et al., 2019).

The Adam optimizer (Kingma & Ba, 2014) is a similar approach to Adadelta, in that it displays the ability to handle non-stationary and noisy data for EEG (Hefron, Borghetti, Schubert Kabban, et al., 2018). SGD updates parameters for each training trial X^{ni} and its hot label vector Y^{ni}, SGD has been reported be faster than its equivalent on Batch Gradient Descent that performs parameter updates based on a large portion of the datasets. However, as a trade-off, SGD can display high variance in objective stability during learning (Ruder, 2016).

Adadelta (Zeiler, 2012) is an extension of Adagrad (Duchi et al., 2011). This optimization algorithm models the geometry of the data to prioritize the occurrence of informative and rare features, in contrast to stochastic gradient approaches that are naïve to data characteristics (Zeiler, 2012). In the context of EEG, Adadelta can influence the deep learning model to discover rare and predictive features in EEG data, while mitigating emphasis on predictive and common potential features such as noise (e.g. frontal noise from blinks, electrical noise from muscles). Even after cleaning, artefacts may still be present in the signal data; for mental workload estimation, this additional signal information may be predictive of mental workload (Borghini et al., 2014; Veltman & Gaillard, 1998; Wilson, 2002).

Nadam (Dozat, 2016) is a variant of Adam that uses Nesterov Momentum. Adam relies on traditional momentum functions, which influence model direction during training, and RMSprop, which adjusts parameters to influence directionality. (Sutskever et al., 2013) showed that Nesterov Momentum could theoretically and empirically affect performance outcomes. Neither in our literature search nor in a systematic review of the field (Roy, Banville, et al., 2019). However, we decided to include it in our benchmarking experiment as it is a variant of Adam, a common optimizer in EEG decoding research, which may prove useful for our goal.

3.2.2.2 Trial wise training and window size

Data window size can have a significant impact on accuracy. Schirrmeister et al. (2017) showed improved accuracy when the data window size was matched to the task trial window for motor movement classification, and it has also been shown that increasing temporal sequence length can positively impact accuracy (Hefron, Borghetti, Schubert Kabban, et al., 2018). However, these results contrast unfavorably with research on emotion recognition, which showed that increasing window size can reduce the accuracy and generalisability of deep learning approaches to novel data due to electrical signal nonstationarity (AlZoubi et al., 2009). Furthermore, conflicting evidence on increasing the data window size is also presented in Almogbel et al. (2018), Almogbel, Dang and Kameyama (2019). It is important to note that in classification tasks with EEG data, the data window size is similar to model hyperparameters and can impact classification accuracy (Saadati et al., 2020b).

The choice of window size in the case of neurological data can be based on domain knowledge of the brain-related phenomena (short responses such as ERPs or longer responses such as mental states) and previous empirical evidence. An ensemble of neural networks with full trial windows of an n-back task (a popular task for inducing variations in mental workload) EEG data as inputs were implemented by Kim et al. (2014). In a related n-back study, Saadati et al. (2020a) tested window sizes between 2s and 5s. They observed the best results when classifying workload using a 3s window. In our approach, we compare two data window sizes of an entire trial window: 3s from -100ms to 2900ms and a half-trial window: from -100ms to 1400ms in order to replicate these results, avoid increased training time and try to maximise performance.

3.2.2.3 Dropout

Dropout is a regularization technique introduced by (Srivastava et al., 2014) that randomly deactivates neurons during model training. Neuron activation is temporarily removed during the forward pass, and the weights are not updated during the backward pass. This reduces overfitting, improves model sparsity, and potentially improves feature quality by reducing co-adaptation. The trade-off, however, is that dropout can increase training time (Srivastava et al., 2014) by increasing in the number of learning epochs and the learning rate.

Dropout has been shown to improve performance in deep learning approaches with EEG (Plis et al., 2014) and has been used in several studies for workload prediction from EEG (Hefron et al., 2017; Jiao et al., 2018; Kuanar, Athitsos, Pradhan, Mishra, & Rao, 2018; Qiao & Bi, 2020; Saadati et al., 2020a). In this study, we test the effect of the dropout rate on hidden layers (i.e., a fixed percentage of inputs are randomly deactivated during a training cycle) for mental workload estimation directly from EEG signal data. The dropout for the visible layers was set to zero to maintain an unaltered and fixed representation of the signal at the input nodes of the models. Qiao and Bi (2020) systematically tested different levels of dropout between 0.1 and 0.9 for deep neural networks and showed peak performance with 50% of neurons randomly deactivated. However, these results are based on estimating mental workload from a spatial and spectral representation of EEG in images. In this research, we test the effect of dropout on the performance of end-to-end neural networks directly from the sensor signal.

3.2.2.4 Activation functions

Activation functions determine the output of a neural network and whether a neuron should be activated based on its inputs. In the case of EEG, Schirrmeister et al. (2017) argue that the choice of activation functions can be sensitive to specific features. Unfortunately, our literature review showed that the activation functions used within models were not reliably reported, reducing the reproducibility of some results. The Rectified Linear Unit (ReLU) is a commonly used function for mental workload decoding with deep learning (Mohammad A Almogbel et al., 2018; Almogbel, Dang, & Kameyama, 2019; Gao et al., 2019; Hefron, Borghetti, Schubert Kabban, et al., 2018; Jiao et al., 2018; Kuanar, Athitsos, Pradhan, Mishra, & Rao, 2018; Kwak et al., 2019; Qayyum, Khan, et al., 2018; Qiao & Bi, 2020; Saadati et al., 2020a). The ReLu activation function has been shown to reduce the training time of very deep neural networks (Glorot et al., 2011) and to improve their classification performance (Dahl et al., 2013). Another activation function is the Exponential Linear Unit (ELU) (Saadati et al., 2020a; Schirrmeister et al., 2017). The ELU has been used within CNNs to classify mental workload Saadati et al. (2020a, 2020b) and for motor activity (Schirrmeister et al., 2017), and both studies reported increased accuracy. Several research approaches utilized other non-linear activation functions, such as the sigmoid function (Fukuda et al., 2019; Hefron et al., 2017; Tao et al., 2019; Yin & Zhang, 2017a, 2018; Yin et al., 2019) or Parametric ReLU (PReLU) (Liu & Liu, 2017). However, they do not provide a rationale nor benchmark results of their application.

3.2.2.5 Design choices

Building on the above literature review, we derived a number of architectural design choices and then systematically tested each one to determine their influence on performance when employed for estimating mental workload using EEG data. Table 16 encapsulates the design choices derived from the review, encompassing aspects such as the optimizer, the training window, the activation function, and the dropout rates. These design decisions are evaluated within the context of the architectures chosen following the benchmarking phase.

Table 16

Design Choices and Justification

Design Choice	Variant	Justification
Optimizers	Adam, SGD, Adadelta, Nadam	Rapid convergence and stability Features identification
Training window	Full trial windows, half trial windows	The varying length of windows can impact the performance as shown
Activation function	ReLU and eLU	Relu as shown to be faster in the hidden layers for deep neural networks, eLU shown to perform well with residual gates in EEG learning tasks
Dropout	From 0.5 to 0.2 on hidden layers	Coadaptation Force the model to learn generalizable features instead of focusing on highly predictive one Performance

Note. Variant column displays the parameters used during model benchmarking. Design choice evaluations were performed on the selected models after architecture selection.

We also implemented several default design choices, such as categorical cross-entropy as the loss function, automatically reducing the learning rate by a factor of 0.5 when the training loss did not improve with a patience of 50 epochs. In addition, we use early stopping with a model checkpoint.

3.2.3 The n-back task for mental workload estimation

Successful training of any machine learning classification method requires an objective ground truth to provide annotated examples of the target signal for classification. In the case of the research presented in this manuscript, a key requirement is the reliable induction of varying levels of mental workload. After reviewing the literature, we selected the n-back task based on the growing empirical evidence that this task can reliably induce and predict individual differences in higher cognitive functions such as low, medium and high mental load (Jaeggi et al., 2010).

The n-back has been used in many studies to induce different levels of mental workload and is widely accepted as a training task for mental workload estimation (Baldwin & Penaranda, 2012; Hefron, Borghetti, Schubert Kabban, et al., 2018; Kim et al., 2014; Kuanar, Athitsos, Pradhan, Mishra, & Rao, 2018; Saadati et al., 2020b). The n-back has been used in EEG studies to estimate workload by deriving features from EEG signal data in the time domain (e.g. ERPs) (Brouwer et al., 2012; Saadati et al., 2020a, 2020b) and the frequency domain (e.g., spectral power) (Brouwer et al., 2012; Hefron, Borghetti, Kabban, et al., 2018). Brouwer et al. (2012) and Berka et al. (2007) both presented evidence that mental workload induced by n-back can be successfully classified during single trials. Moreover, Baldwin and Penaranda (2012) and Yin and Zhang (2018) demonstrated that an artificial neural network trained for mental workload estimation with EEG could be transferred to classify mental workload during the performance of other tasks. Two properties of the n-back task help explain the generalizability of the inference: both the perceptual and motor demands remain constant across difficulty levels of the task (Grimes et al., 2008). Furthermore, it has been posited that the n-back task manipulates mental workload through additive working memory load, and not a reaction to a conceptually different stimulus (Grimes et al., 2008). Thus, we selected this paradigm to elicit different level of mental workload.

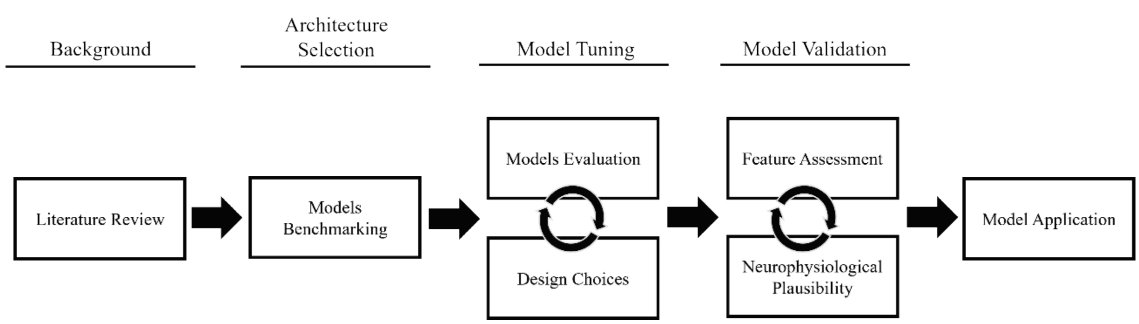
3.2.4 Methodological framework

Our methodology builds on the framework proposed by Kohoutová et al. (2020) on the use of machine learning for decoding and neuroimaging. This methodological framework is fundamental, given that end-to-end deep learning approaches present unique challenges related to EEG signal properties. For example, due to the use of raw sensor-level data, it is prone to confounding, non-random noise capture, and signal non-stationarity. Our methodology follows four steps prior to model application: (1) background, (2) architecture selection, (3) model tuning, and (4) model validation (Figure 14).

Firstly, a literature review was conducted to survey the literature to inform the research process and identify gaps. This activity also informs the phases of selecting the architecture, tuning the model and validating the model. In addition, this activity helps us generate architectural design decisions that can impact performance. Moreover, it also provides empirical evidence on the neurophysiological response of interest, which supports the plausibility assessment of the features.

Figure 14

Methodological Framework



Secondly, a benchmarking procedure is developed to select the best performing architecture. The aim is to select baseline models and specific architectures for which evaluations are deemed to contribute to answering the problem. Finally, the best performing models are selected for further tuning.

Thirdly, we tune model architectures based on design choices and evaluate their impact on performance. In line with Schirrmeister et al. (2017), design choices are motivated based on the impact on task performance and generalizability in the study context.

Fourthly, as previously discussed, caution must be exercised when applying end-to-end processes to EEG data without prior knowledge of the cognitive construct or the resulting dataset; the properties of the EEG signal and model tuning design choices will lack a plausible inference model. Predictive artifacts can confound and affect the performance and accuracy (either positively or negatively) of the model, depending on the task. Moreover, care must also be taken to ensure that only brain response signals are used for classification. In the case of end-to-end approaches, features are learned automatically, as opposed to engineered features, which can result in complex features to interpret (Nguyen et al., 2015). Thus, end-to-end approaches need further validation to ensure that the performance is not due to the model fitting noise or artifact signal (Lawhern et al., 2018). The model validation phase aims to ensure that the trained models learn from relevant neurological signals, not from prediction artefacts or confounds. The tuned models are evaluated for neurophysiological plausibility, and features are compared with EEG data and MW decoding evidence. This validation phase ensures that there is converging evidence from the neuroscience literature.

3.3 Materials and Experiment Design

3.3.1 Experimental environment

The study took place at an aerospace company's training facility. Data were collected from within a high-fidelity prototype of a flight simulation presented through the CAE Medallion MR e-Series visual system. The visual display consists of partial sphere screen spanning 200°, measuring 1m in radius and 1.5m in height, providing a viewable surface area of 9.42m. Participants were situated within an aircraft cockpit designed to scale with a replicated flight control system. The environment offered the operator an ecologically valid and realistic flight simulation.

During the experiment, participants performed two experimental tasks: a synthetic task consisting of an n-back designed to manipulate mental workload, and an ecologically valid flight task designed to induce different levels of mental workload by manipulating maneuver complexity. The basis of the experimental design is that the manipulation of difficulty within the flight task mirrors the MW induced by the synthetic task (i.e., n-back difficulty) while maintaining a valid flight training scenario.

3.3.1.1 Synthetic task – n-back

During the n-back task, the participant is asked to monitor a series of stimuli to memorize and compare unique items. Participants are required to respond through a keypress when the presented stimulus is the same as one presented in n trials before, where n is a parameter used to induce different levels of mental workload through the manipulation of working memory capacity. In this instance, the left arrow key was used for non-target letters and the right arrow key for target letters. We used a verbal stimulus (i.e., letters) and identity recognition (i.e., the same letter as presented n trials previously) design. To obtain a granular classification of working memory level, we opted to increment the number of iterations of n from 0 to 3, giving four difficulty levels.

Each iteration of n is composed of 40 trials. Before each iteration, participants were allowed ten practice trials. The stimuli for each trial within an iteration of n were presented in a pseudo-randomized order for targets (35%) and non-targets (65%), for 500 milliseconds followed by a 3000-millisecond interstimulus interval in which a cross-shaped fixation character was displayed. The task was implemented using E-Prime 3.0 (Psychology Software Tools, Pittsburgh, PA).

3.3.1.2 Naturalistic task – flight simulation maneuvers

The simulated flight training scenario was developed to mimic and induce similar levels of mental workload as that produced by the n-back task. We followed a single task paradigm to induce three levels of MW through maneuver complexity. In this instance, three blocks of flight maneuvers represent low, moderate, and high complexity (Table 17). Each maneuver was coupled with appropriate flight actions and directed by a professional flight instructor, who logged each participant's performance, set the pace of the training scenario, and signaled the end of each maneuver. In this way, each flight maneuver was aligned with flight telemetry allowing for synchronization flags to be produced and associated with EEG signal data. A training maneuver (a) was utilized to introduce a simple "fly straight" and level maneuver where participants were asked to maintain heading, and banking angle, to familiarize participants with the simulator and the format of the training scenario and altitude.

Table 17

Flight complexity blocks and maneuvers

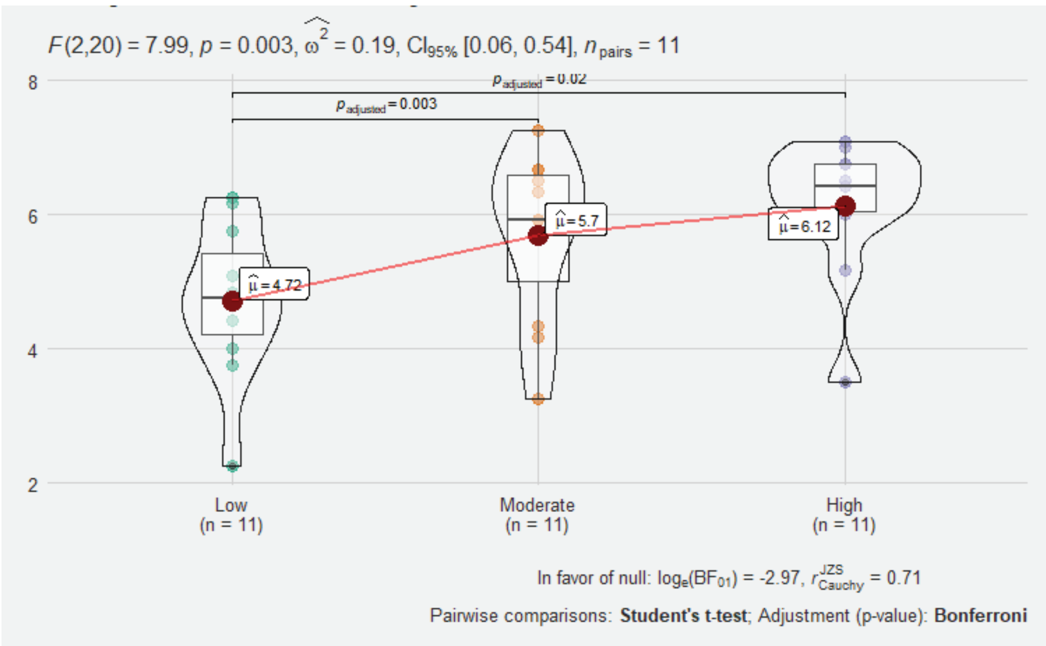
Low	Moderate	High
(b) Speed change	(f) Speed and Altitude	(j) Altitude, Speed, and Heading
(c) Altitude change	(g) Heading and Banking	(k) Altitude, Speed, Heading, and Banking
(d) Heading change	(h) Altitude and Heading	(l) Vertical Loop (altitude, heading, banking)
(e) Banking change	(i) Speed and Banking	

Note. Single task paradigm for three levels of complexity: low, moderate, and high.

In order to assess a participant's perceived assessment of mental workload and as a manipulation check, participants were asked to complete the NASA RAW-TLX (Hart, 2006; Hart & Staveland, 1988). This approach has been previously reported as an effective method of assessing subjective changes in MW in response to task complexity. Descriptive statistics show that the mean subjective mental workload for the low complexity maneuvers is the lowest with 4.72 (±1.15), followed by 5.70 (±1.25) for the moderate complexity, and 6.12 (±1.03) for the high complexity maneuvers. We performed a single factor, within-subject repeated-measures ANOVA to determine the statistical significance between perceived mental workload and each maneuver block. Assumptions of normality and sphericity (Mauchly's test $p = .093$) were met. The difference between means was statistically significant, $F(2,20) = 7.995$, $p = .003$. Post hoc analysis using Student's t-test and Bonferroni adjusted alpha levels indicated that the average perceived mental workload was significantly lower in the low complexity condition than in the moderate condition, p = .003, and the high condition, p =0.02 (Figure 15).

Figure 15

Violin and box plot of manipulation check for the simulator task



Note. X axis represents the complexity level, the Y axis the perceived mental workload measured with the NASA RAW-TLX

3.3.2 Participants

Eleven participants (3 females) with a mean age of 34.56 years (SD = 8.97), all novice airplane pilots, participated in the study. Novice pilots were selected due to the choice of the naturalistic task selected, see next section – experimental environment. Participants were screened on the basis of good health and normal to corrected to normal vision. The study received ethical approval from the institution ethics review committee regarding data security (ethic certificate number: 2020-3559). Ethical approval was also obtained through an internal review by the hosting aerospace company, and participants were recruited from within the company.

Each participant signed consent following our institution ethical committee guidelines. Participants were further instructed that they could depart at any time during the experimental procedure. These instructions and subsequent consent were provided a priori, privately through communication with the primary researcher rather than through company channels. Participants were also instructed to avoid alcohol and caffeine before the experiment.

3.4 Data Processing, Learning Task and Inputs

3.4.1 Data processing

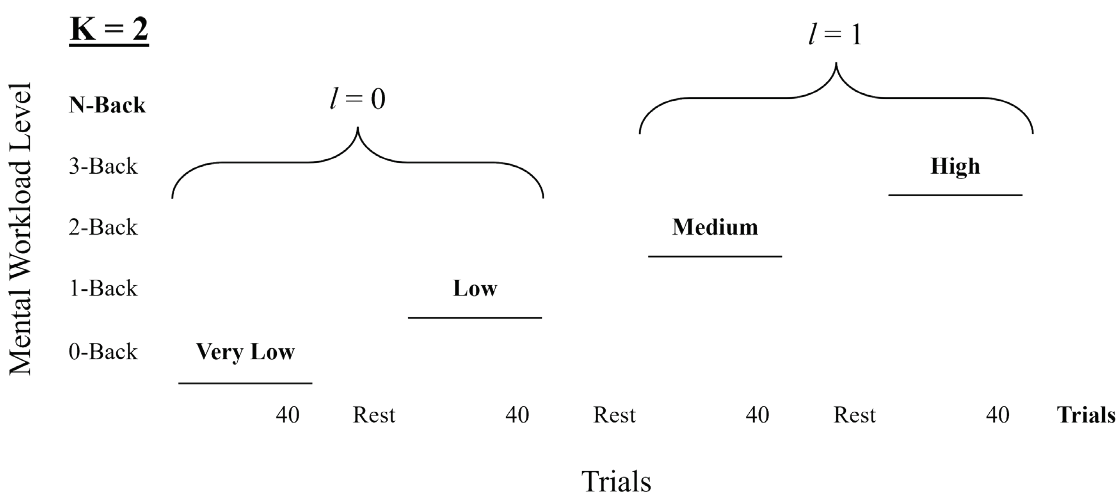
EEG signal was recorded using a 32-channel BrainVision (Morrisville, USA) headset following the standard 10–20 montage. The signal was first referenced to Fz, then filtered using a bandpass 1-40 Hz Butterworth 2nd order IIR filter. Artifacts were removed using blind source separation by independent component analysis (extended infomax). During the signal quality assessment, data for two subjects were rejected: one due to poor signal quality and the other due to issues with event synchronization. The filtered EEG signal data was then down sampled to 500 Hz (from 1000Hz), then segmented by task, and split into n=160 epochs per participant for a total of 1440 epochs. We chose full-trial windows of 3 seconds (-100ms to 2900ms) for each epoch to provide sufficient data to inform design decisions regarding to the effect of different window sizes on the deep learning modeling process.

3.4.2 Problem, learning task, and inputs

The modeling goal in this instance is to train and test a number of discriminative deep learning models f on recorded EEG data from the *n*-back task such that: $f\colon \mathbb{R}^{S \times T} \rightarrow \{K\}$ where K is a vector of classes, S represents the cleaned sensor-level signal, and T the time steps of the recorded signal. For instance, in this study, S is composed of 32 channels, T represents 1500 time steps at sampling rate of 500 Hz for a window of 3 seconds. The task is to classify brain signals into a probability distribution of *n* class labels. The learning problem consists of a task with K classes of j length equivalent to a 2-class problem. We derive K=2 classes, by concatenating the n-back data per level were $n = 0$ and $n = 1$, represent low workload, and $n = 2$ and $n = 3$ represent high workload (see Figure 16). We took this approach for two reasons: discriminative features in sensor signals can be challenging when algorithms are trained to differentiate between mental workload levels where the variance between task difficulty is low and given the relatively low number of training examples.

Figure 16

Class labels for mental workload estimation



Note. Class aggregation where K = 2 to create a two-class problem composed of a binary aggregate of the original 4 task difficulty condition. Each condition is composed of 40 trials per participant giving 80 trials per class.

Each training input was designed as a subject dependent dataset i composed of multivariate time series matrixes representing a trial j. More precisely, we define j as an S-dimensional multivariate time-series as $X^{ji} = [X_1^{ji}, X_2^{ji}, \dots, X_S^{ji}]$ composed of S unique univariate time-series, $X_s \epsilon \mathbb{R}^T$, where $X_s = [x_1, x_2, \dots, x_T]$ with a length of X equal to a finite number of real values S and T represents the cleaned sensor-level signal over time. Then, for a given subject i, we have a dataset $D^i =$

$\{(X^{1i}, Y^{1i}), \ldots, (X^{ni}, Y^{ni})\}$, where for a pair (X^{ni}, Y^{ni}), X^{ni} is a multivariate time series, $Y^{ni}$is the related one-hot label vector of the trial, and Ni represents all the recorded trials j for subject i. For a task containing K classes, Y^{ni} is a vector of length K with $j \epsilon [1, K]$.

3.4.3 Architecture selection and baselines

3.4.3.1 Models

Informed by the literature, we identified a gap in the knowledge base related to mental workload classification using deep learning approaches as end-to-end processes to create internally generated discriminative features. To address this gap and using previously reported deep learning methods for engineered features and domain-specific techniques for mental workload classification as a starting point, we identified several deep learning models with the potential for end-to-end discriminative feature generation. We selected elven models for testing and evaluation based on the uniqueness of their architecture, their centrality in the literature, and the recent advances in the state-of-the-art of end-to-end deep learning models for time series classification (Fawaz et al., 2019; Lawhern et al., 2018; Schirrmeister et al., 2017): a Multi-Layer Perception (MLP) (Wang et al., 2017), a Fully Convolutional Neural Network (FCN) (Wang et al., 2017), a Multi-Channel Deep Convolutional Neural Network (MCDCNN) (Zheng et al., 2016), a Residual Network (ResNet) (Wang et al., 2017), an Encoder (Encoder), and a Recurrent Neural Network with Long short-term memory (RNN+LSTM), a Deep Convolutional Network (Deep Conv Net) (Schirrmeister et al., 2017), a Shallow Convolutional Network (Shallow Conv Net) (Schirrmeister et al., 2017), and EEGNet (Lawhern et al., 2018). The initial architectures are implemented as describe in the manuscripts. Finally, we added a simple perceptron (PPN) as a baseline deep learning model. Table 18 provides a description of the architecture of the benchmarked models.

The selected models are unique architectural variants of convolutional neural networks reported to possess advantageous characteristics and interesting performance when applied to multivariate time series datasets. These models learn hierarchical representations directly from raw input signals and provide a probabilistic inference of the class prediction through successive non-linear transformations.

Table 18

Architectures for the benchmarked models

	PPN	CNN	FCN	EEGNet	Deep Conv Net	MCDCNN	MLP	ResNet	RNN-LSTM	Encoder	Shallow Conv Net
Layers	1	3	5	4	5	4	4	11	4	5	2
Convolution	0	2	4	3	4	2	0	9	0	3	1
Invariance	0	2	4	3	4	2	0	10	0	4	1
Normalize	None	None	Batch	Batch	Batch	None	None	Batch	None	Instance	Batch
Pooling	None	Avg	None	Avg	Max	Max	None	None	None	Max	Avg
Feature	FC	Conv	GAP	Conv	Conv	FC	FC	GAP	FC	Att	Conv
Activation Function	Sigmoid	Sigmoid	ReLU	ReLu	eLU	ReLU	ReLU	ReLU	Sigmoid	PReLU	Cust
Regularization	Dropout	None	None	Dropout	Dropout	None	Dropout	None	Dropout	Dropout	Dropout
Optimization Algorithm	Adam	Adam	Adam	SGD	Adam	SGD	AdaDelta	Adam	Adam	Adam	Adam
Loss function	Entropy	MSE	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy
Epochs	2000	2000	2000	2000	1500	120	5000	1500	10	100	1500
Learning Rate	.01	.001	.001	.001	.001	.01	.01	.001	.001	.00001	.001

3.4.3.1 Baselines

We use a varied set of baseline models to compare the selected end-to-end deep learning models (Table 19) with both simple and state-of-the art baselines. Nevertheless, all the baseline models could be classified as end-to-end approaches, given their minimal reliance on hand-engineered features, thereby ensuring a fair comparison. First, a simple end-to-end perceptron is implemented to have a baseline in deep learning. Secondly, we use non-deep learning multiple classification techniques known to perform well in EEG data. The simplest model is a linear regression on sensor data (LR), then we use preprocessing techniques with an estimation of covariance matrix (COV) and covariance matrix with Xdawn spatial filtering (Xdawn) before applying LR (Rivet et al., 2009). As a fourth baseline, we transform the EEG signal with a Principal component analysis (PCA) and signal decomposition using the Common Spatial Patterns (CSP) before applying a shrinkage Linear Discriminant Analysis (sLDA) as it is deemed more robust than classic LDA (Lotte et al., 2018). Finally, we use COV followed by Riemannian Minimum Distance to Mean (RMDM) as it is known to be one of the state of the art techniques in EEG decoding (Appriou et al., 2018; Lotte et al., 2018).

Table 19

Baseline models and implementations

Baseline Model	Implementation
COV+LR	Estimation of covariance matrix Data are vectorized (across time and channels) Features standardized by removing the mean and scaling to unit variance (normalize features across trials) Logistic Regression on covariance matrix
COV+RMDM	Estimation of covariance matrix Riemannian Minimum Distance to Mean (RMDM)
LR	Data are vectorized (across time and channels) Features standardized by removing the mean and scaling to unit variance (normalize features across trials) Logistic Regression on sensor data
PCA+CSP+LDA	Principal component analysis (PCA) Signal decomposition using the Common Spatial Patterns (CSP) shrinkage Linear Discriminant Analysis (sLDA)
Xdawn+LR	Covariance matrix with Xdawn spatial filtering (Xdawn) Data are vectorized (across time and channels)

	Features standardized by removing the mean and scaling to unit variance (normalize features across trials) Logistic Regression on covariance matrix
PPN	Architecture defined with the deep learning models

3.4.4 Statistical testing

Benchmarked models are evaluated by calculating three standard performance metrics, accuracy, precision, and recall. The $Accuracy = \frac{TP+TN}{N}$ calculates the total of correctly classified classes for the total number of predictions, which is equal to $\frac{TP+TN}{TP+TN+FP+FN}$ in a binary classification where TP represents the total true positive, TN represents the total true negative, FP the false positive, and FN the false-negative classifications, respectively. We also calculate precision and recall. $Precision = \frac{TP}{TP+FP}$ represents the proportion of correctly predicted positive instances (true positives) out of all instances predicted as positive. $Recall = \frac{TP}{TP+FN}$ also known as sensitivity or true positive rate, measures the proportion of correctly predicted positive instances (true positives) out of all actual positive instances. Statistical tests are applied to the accuracy scores. We acknowledge that due to the balanced dataset of each class, combined with the utilization of macro-average recall, may result in similar accuracy and recall scores.

We compute the statistical significance of the mean difference in performance between deep learning models to systematically test model architectures and design choices. We utilized Wilcoxon signed-rank tests, as suggested by (Benavoli et al., 2016). Additionally, we applied the Benjamini-Hochberg procedure for false-discovery-rate correction with a nominal α = 0.05 and a false discovery rate of 10% (Benjamini & Hochberg, 1995). A critical difference diagram to visually present post-hoc significant and non-significant pairwise comparisons of the mean ranks. A horizontal line between the ordered tested model shows a non-significant difference.

3.4.5 Training hardware and software

The deep learning models were trained using compute hardware composed of an Intel® Core™ i7-8700K @ 3.70GHz CPU, 32 GB of random-access memory (RAM), and one

NVidia Titan Xp GPU. Models were implemented using the open-source library Keras 2.3.0 (Chollet, 2015), a high-level neural network library on Python built over TensorFlow 2.14 (Abadi et al., 2015).

3.5 Architecture Selection

This section is organized to present the performance assessment of the baselines, the eleven DL models trained and tested using the benchmark task (n-back), each model was trained over each participant using 5-fold cross validation. Details about the statistical procedure are available in appendix A2.4.

3.5.1 Baselines

Based on the average accuracy across all participant's datasets and 5-fold cross validation, the results show (Table 20) that the COV+LR and COV+RMDM baseline models outperform the other models when classifying mental workload where K = 2 (high or low MW) from the EEG. COV+LR reports a mean accuracy of .874 (± .088) and COV+RMDM .884 (± .098). In this instance, a low SD is representative of model stability across participant data.

Table 20

Baseline Models Performance Results

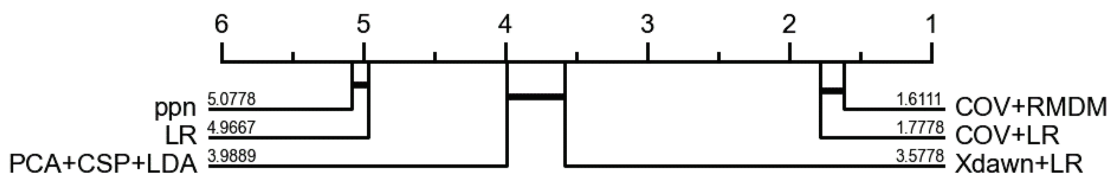
Models	Accuracy		Precision		Recall	
COV+LR	.874	±.088	.880	±.086	.874	±.088
COV+RMDM	.884	±.098	.889	±.094	.884	±.098
LR	.601	±.091	.605	±.093	.601	±.091
PCA+CSP+LDA	.678	±.127	.684	±.128	.678	±.127
Xdawn+LR	.708	±.111	.711	±.110	.708	±.111
PPN	.590	±.090	.593	±.092	.590	±.090

Note. K = 2. The results are the average of over five cross-validations. Metrics are expressed as average ± standard deviation.

A Wilcoxon Signed-Ranks test was conducted to determine the classification accuracy of the COV-RMDM (mean rank = 1.611) and COV+LR (mean rank = 1.778) models in comparison to the other trained models (Figure 17). The results indicated that both models demonstrated a significantly higher classification accuracy than the other models. However, their performance was similar with no significant difference between them ($Z = 458.5$, $p = 0.504$). COV-RMDM was subsequently used as a baseline for the subsequent benchmarking process.

Figure 17

Critical plot difference for the baseline models



Note. Critical Plot Difference based on Wilcoxon Signed-Rank test with Benjamini-Hochberg procedure for false-discovery-rate correction for the Baseline models. A straight horizontal line between the models shows a nonsignificant difference in post-hoc pairwise comparisons. Values show the average rank of the model.

3.5.2 Models benchmarking

Based on the average accuracy across all participant's datasets and 5-fold cross validation procedure, Table 21 presents the models' performance metrics and the average training duration. The results show that the FCN and ResNet models outperform the other models and baseline when classifying mental workload. FCN reports a mean accuracy of .939 (± .052) and .921 (± .073) for the Resnet, followed by the EEGNet (mean = .873 ± .101) and the ShallowConvNet (mean = .846, ± .106). All end-to-end deep learning models were compared between each other and against the best performing baseline (Figure 5), i.e., COV+RMDM.

Table 21

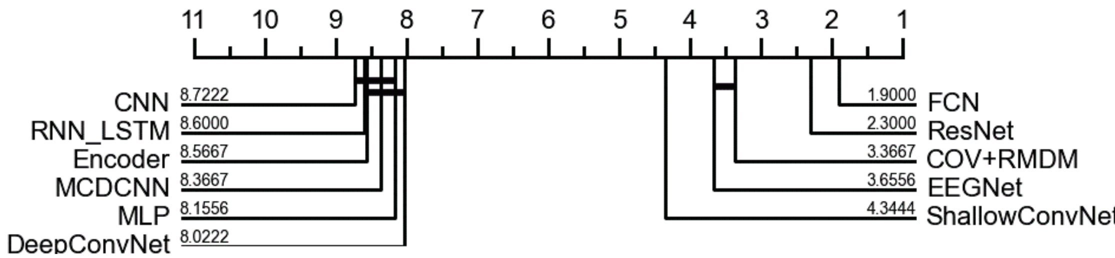
Descriptive Statistics Benchmarked Models

Models	Average of accuracy		Average of precision		Average of recall		Average of duration (sec)	
CNN	.588	±.091	.590	±.092	.588	±.091	187.787	±6.464
Encoder	.598	±.100	.602	±.103	.598	±.100	63.288	±1.685
DeepConvNet	.635	±.099	.648	±.112	.635	±.099	550.378	±4.358
FCN	.939	±.052	.942	±.049	.939	±.052	316.535	±11.131
EEGNet	.873	±.101	.879	±.100	.873	±.101	232.858	±4.492
MCDCNN	.611	±.114	.611	±.187	.611	±.114	38.527	±1.816
MLP	.615	±.069	.620	±.071	.615	±.069	783.523	±23.975
ResNet	.921	±.073	.925	±.070	.921	±.073	515.206	±14.190
RNN LSTM	.603	±.104	.606	±.106	.603	±.104	581.033	±24.224
ShallowConvNet	.846	±.106	.851	±.105	.846	±.106	181.242	±2.398

A statistical comparison of the models using a Wilcoxon signed-rank test indicates that FCN (mean rank = 1.900) and ResNet (mean rank = 2.300) present classification accuracy superior to the baseline (mean rank = 3.367), EEGNet (mean rank = 3.656) and the other trained models (Figures 18). The Wilcoxon signed-ranks test indicated that FCN is significantly more accurate than COV+RMDM ($Z = 135$, $p < 0.000$) and ResNet ($Z = 325.5$, $p = 0.027$). ResNet is significantly more accurate than COV+RMDM ($Z = 220.5$, $p = 0.001$). COV+RMDM showed no significant difference in accuracy than EEGNet ($Z = 452.5$, $p = 0.461$). Based on those results, FCN and ResNet were selected for further investigation.

Figure 18

Critical plot diagram of the deep learning models against baseline



Note. Baseline = COV + RMDM.

3.6 Model Tuning

3.6.1 Optimization algorithm

In order to optimize the training of our model, we tested four different optimization algorithms. As discussed in the literature review, optimization techniques aim to reduce the difference between the training and test errors of the model by minimizing the cost function with respect to its parameters. The optimization algorithms were selected based on their internal mechanisms and previous research in the relevant field.

Our objective was to choose optimizers that minimize the model training time while ensuring stable convergence and taking into account the unique properties of each optimizer. To achieve this, we tested four different optimizers: Stochastic Gradient Descent (SGD), AdaDelta, Adam, and Nadam. The descriptive statistics revealed that using any of these optimizers (Adam, AdaDelta, Nadam or SGD) produced stable classification accuracies for both FCN and ResNet (Table 22).

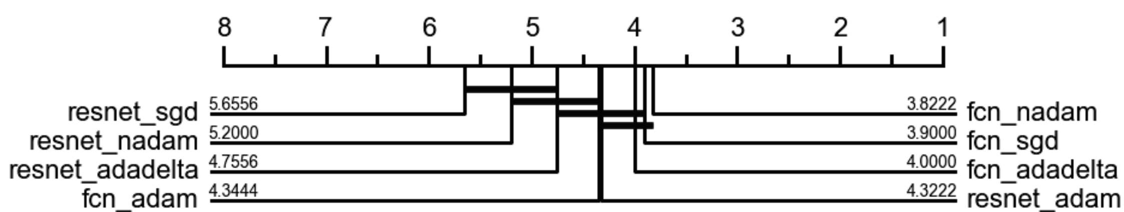
Table 22

Descriptive statistics for the optimizer choice

Model	Optimizer	Accuracy		Precision		Recall		Duration	
FCN	Adadelta	.937	±.057	.939	±.056	.937	±.051	311.222	±3.109
	Adam	.935	±.057	.938	±.054	.935	±.057	311.178	±4.941
	Nadam	.941	±.054	.944	±.051	.941	±.054	310.546	±3.325
	SGD	.941	±.044	.944	±.042	.941	±.044	295.880	±3.982
Resnet	Adadelta	.926	±.065	.929	±.062	.926	±.065	520.839	±5.147
	Adam	.926	±.067	.930	±.064	.926	±.067	515.730	±4.602
	Nadam	.915	±.079	.918	±.077	.915	±.079	537.711	±6.076
	SGD	.913	±.069	.919	±.063	.913	±.069	479.178	±4.863

Figure 19

Critical Difference Diagram for Optimizers



For FCN, SGD and Nadam showed similar accuracy with .941 (±.044) and .941 (±.054), respectively. For Resnet, the best performing optimizer was Adam with an accuracy of .926 (±0.067). As shown by Figure 19, a statistical comparison of the models using a Wilcoxon signed-rank test indicates that Nadam (mean rank = 3.822) and SGD (mean rank = 3.900) were the two best performing optimizers for FCN. The Wilcoxon signed-ranks test indicated no significant difference in the mean accuracy between the two (Z = 514, p = 0.967). For ResNet, the results show not statistical difference between Adam (mean rank = 4.322) and Adadelta (mean rank = 4.756), Z = 514, p = .968). Based on these results, we selected Nadam as the optimizer of choice for FCN; Adam was retained for ResNet due to fast convergence and good accuracy.

3.6.2 Dropout rate

In order to analyze the effect of dropout rates on model accuracy, we compared different rates for both the FCN and ResNet models and recorded the results in Table 23. As seen in the literature review, dropout is a regularization technique that randomly deactivates neurons during model training, to reduce overfitting, improve model sparsity, and potentially enhance feature quality by reducing co-adaptation.

The dropout rate of hidden layers and visible layers was set to [0.1, 0.4], and [0.0], respectively. Figure 20 shows the impact of dropout rate on models' accuracies and the critical difference diagrams. For FCN, when using a .4 dropout rate, the average accuracy was reduced to .865 (± .101). For Resnet, the accuracy remained stable; the average accuracy for the .4 dropout rate was .917 (± .069). When using a .2 dropout rate, accuracies remained stable for Resnet, reporting average accuracies of .926 (±.070). Figure 8 shows the accuracy evolution based on the FCN and ResNet dropout rate for each participant and fold. Wilcoxon Signed-Rank tests show a statistical difference between .1 (mean rank = 1.922) and the .2 (mean rank 2.333) rates for FCN ($Z = 90.5$, $p < 0.001$). The ResNet showed no statistical difference when the dropout rate was increased. Based on these results, we selected a 0.2 dropout rate for Resnet for hidden layers to leverage the technique's positive effects on the risk of over-fitting, sparsity improvement, and quality of features learned. For FCN, the dropout rate was set at 0.1.

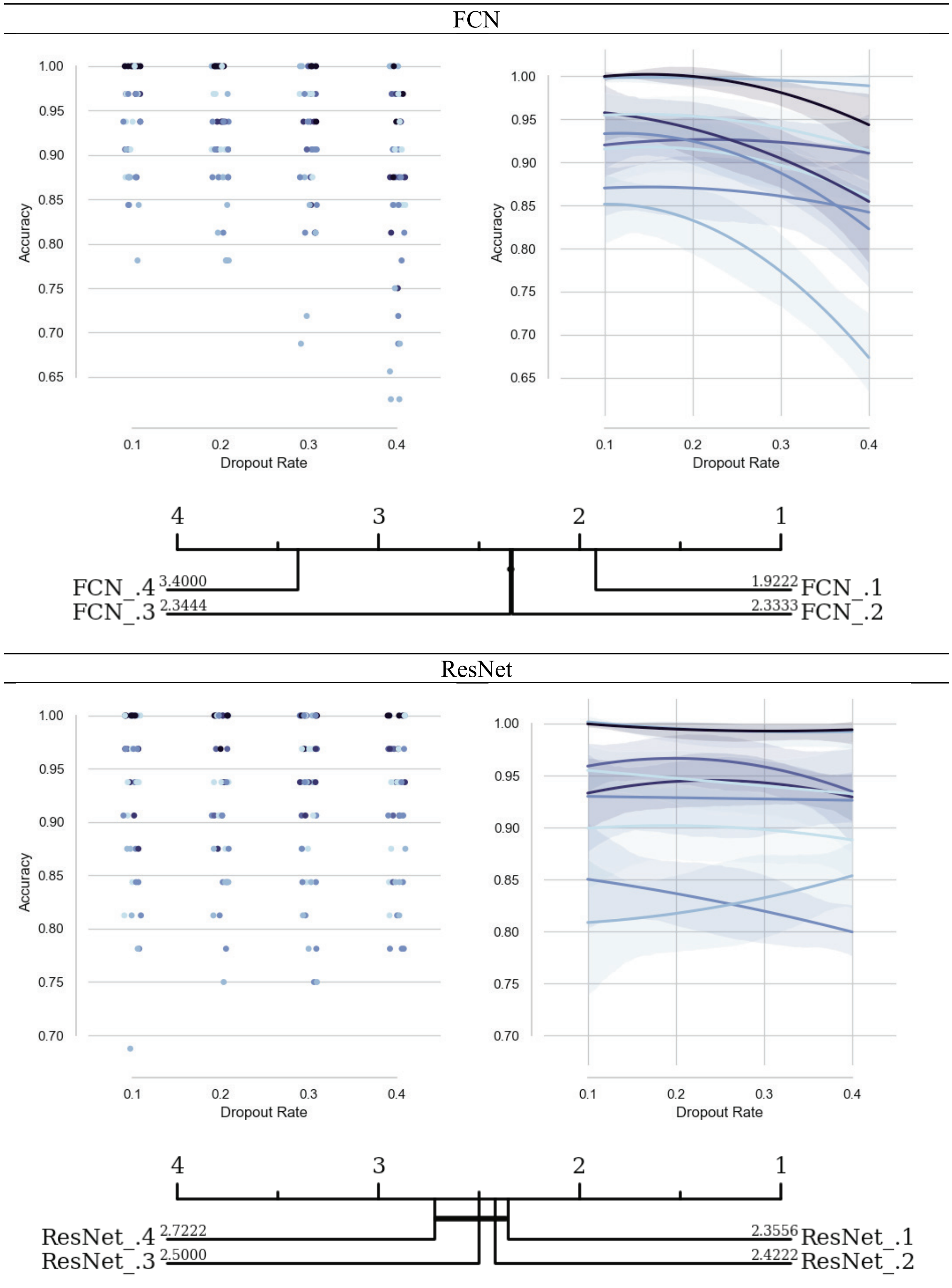
Table 23

Descriptive Statistics for Dropout rate

Model	Dropout	Accuracy		Precision		Recall		Duration	
FCN	.1	.936	±.057	.941	±.052	.936	±.057	329.348	±7.519
	.2	.922	±.066	.931	±.056	.922	±.066	332.489	±6.494
	.3	.915	±.075	.926	±.062	.915	±.075	332.457	±6.466
	.4	.865	±.101	.897	±.066	.865	±.101	331.977	±6.956
ResNet	.1	.926	±.076	.929	±.074	.926	±.076	533.725	±9.526
	.2	.926	±.07	.928	±.068	.926	±.07	533.88	±8.953
	.3	.923	±.076	.929	±.069	.923	±.076	534.69	±9.587
	.4	.917	±.069	.922	±.065	.917	±.069	534.062	±10.283

Figure 20

Impact of dropout rate on accuracy for FCN and ResNet



3.6.3 Activation function

We conducted a comparison of various activation functions for both FCN and ResNet models and recorded their respective impacts on performance in Table 24. Activation functions are responsible for determining the output of a neural network and whether or not a neuron should be activated based on its inputs. As discussed in the literature review, some researchers have suggested that the choice of activation functions for EEG may be sensitive to specific features of the data. For the FCN model, we found that using the Relu activation function yielded an accuracy of .933 (±.062), while using the Elu function resulted in an accuracy of .914 (±.064). As for the ResNet model, we found that using Elu as the activation function resulted in an accuracy of .933 (±.066), while using Relu returned an accuracy of .915 (±.078).

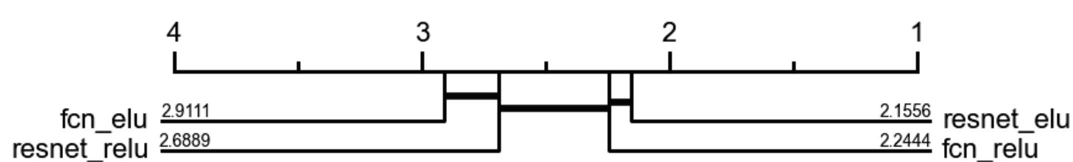
Table 24

Descriptive Statistics for Activation Function performance

Model	Activation Function	Accuracy		Precision		Recall		Duration	
FCN	Elu	.915	±.064	.919	±.06	.915	±.064	326.474	±3.932
	Relu	.933	±.062	.938	±.054	.933	±.062	326.786	±3.773
Resnet	Elu	.933	±.066	.936	±.062	.933	±.066	525.865	±5.987
	Relu	.915	±.078	.922	±.073	.915	±.078	523.997	±4.165

Figure 21

Critical diagram plot for accuracy of activation function



A Wilcoxon Signed-Rank test was conducted to determine the statistical difference in accuracy between the Relu (mean rank = 2.244) and Elu (mean rank = 2.911) activation functions for the FCN model. The results revealed a significant difference in accuracy between these two functions ($Z = 247.5$, $p < .002$). Conversely, we observed a significant difference in accuracy between the Elu (mean rank = 2.156) and Relu (mean rank = 2.6889) activation functions for the ResNet model ($Z = 309$, $p < .017$). Given the statistical analysis results, we have decided to select the Relu activation function for the FCN model and the Elu activation function for the ResNet model.

3.6.4 Window size

To determine the effect of data window size upon training duration and model accuracy, we compared a full trial window of 3s (-100ms to 2900ms) with a 1.5s windows (-100ms to 1400ms) relative to n-back trial onset. In this instance, both FCN and ResNet performed marginally worse than the baseline with the smaller window size (Table 25). The average accuracy for the 3s windows were .033 (±.062) and .933 (±.066) for FCN and ResNet, respectively, compared to .907 (±.08) and .865 (±.114) for FCN and Resnet using 1.5s windows. Furthermore, when using the smaller 1.5s window, ResNet accuracy displays more variance across datasets.

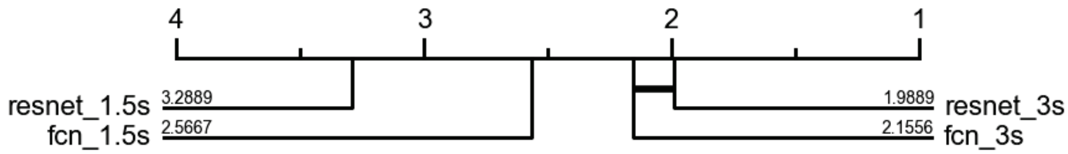
Table 25

Descriptive Statistics for Windows Size of 1.5 seconds and 3 seconds

Model	Accuracy		Precision		Recall		Duration	
fcn_1.5s	.907	±.080	.911	±.078	.907	±.08	80.780	±.684
fcn_3s	.933	±.062	.938	±.054	.933	±.062	326.786	±.773
resnet_1.5s	.865	±.114	.871	±.110	.865	±.114	149.955	±.985
resnet_3s	.933	±.066	.936	±.062	.933	±.066	525.865	±.987

Figure 22

Critical diagram plot for accuracy of window size



We conducted a Wilcoxon Signed-Rank test to examine the statistical difference in accuracy based on the window size used (as shown in Figure 22). Our findings indicate that for the FCN model, a window size of 1.5 seconds (mean rank = 2.567) resulted in significantly inferior accuracy compared to a window size of 3 seconds (mean rank = 2.156) ($Z = 307$, $p = .017$). Similar results were observed for the ResNet model, where a window size of 1.5 seconds (mean rank = 3.289) resulted in significantly inferior accuracy compared to a window size of 3 seconds (mean rank = 1.989) ($Z = 102.5$, $p < .001$). Based on these results, we select a window size of 3 seconds for both the FCN and ResNet models.

3.6.5 Performance replication

In order to determine if the reported accuracies of our chosen models were unique to the data used to train them or generalizable across similar datasets, the two parametrized models were tested on an additional unique n-back dataset composed of 15 different participants collected in a laboratory setting. This time, A 64-channel headset from BrainVision was used to record the EEG. The sensor signal was preprocessed following the same pipeline used for the main dataset. The fifteen participants' brain signal was recorded during the same n-back task implemented with E-PRIME. FCN reported accuracies for this dataset was .909 (±.078). The reported accuracy for ResNet was .900 (±.103).

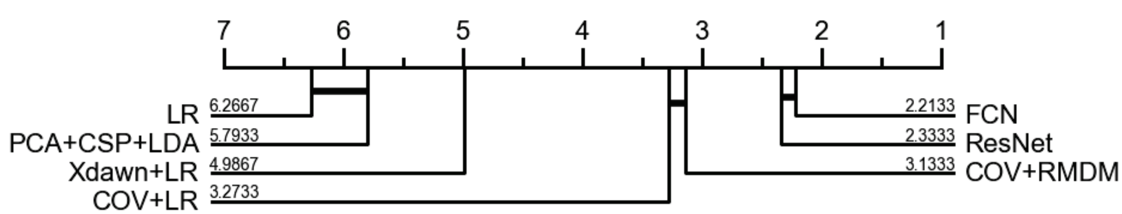
Table 26

Descriptive statistics for the replication procedure

Model	Accuracy		Precision		Recall		Duration	
FCN	.933	±.054	.938	±.050	.933	±.054	326.248	±2.229
ResNet	.917	±.074	.921	±.071	.917	±.074	711.267	±3.293
COV+LR	.861	±.112	.866	±.123	.867	±.111	.477	±.108
COV+RMDM	.861	±.122	.865	±.128	.866	±.162	6.546	±.639
FCN Replication	.909	±.078	.915	±.075	.910	±.078	417.896	±11.914
LR	.635	±.086	.628	±.086	.672	±.138	2.177	±.783
PCA+CSP+LDA	.656	±.099	.661	±.120	.667	±.140	8.387	±1.65
ResNet Replication	.900	±.103	.906	±.098	.900	±.103	607.378	±13.865
Xdawn+LR	.732	±.108	.742	±.117	.722	±.135	3.327	±.639

Figure 23

Critical Diagram Plot for the replication procedure



We conducted a Wilcoxon signed-rank test to determine the accuracy of the FCN and ResNet models on the new dataset. Our findings indicate that there was no significant difference in accuracy between the FCN model (mean rank = 2.213) and the ResNet model (mean rank = 2.333) ($Z = 1375.5$, $p = .793$). However, both end-to-end deep learning architectures still performed significantly better than the baseline COV-RMDM model (mean rank = 3.133). We observed a significant difference in accuracy between the FCN model and the baseline ($Z = 692$, $p < .001$), as well as between the ResNet model and the baseline ($Z = 753.5$, $p < .001$).

3.7 Model Validation

3.7.1 Neurophysiological plausibility assessment

In this section, we conduct the neurophysiological plausibility assessment of the models. The objective is to evaluate the plausibility of the features (Kohoutová et al., 2020) and ensure that our model learns from brain-related signals. We then relate these results to the previous finding in the literature to validate the models. FCN and ResNet learned features directly from EEG signals and can potentially derive correlated information from more than brain signals (such as muscular activities, blinks). EEG has a low-signal-to-noise ratio and captures a variety of information unrelated to the task that can affect the signal more than task-relevant information. Thus, we need converging evidence that the models leverage mental workload relevant information before applying it to the flight EEG data.

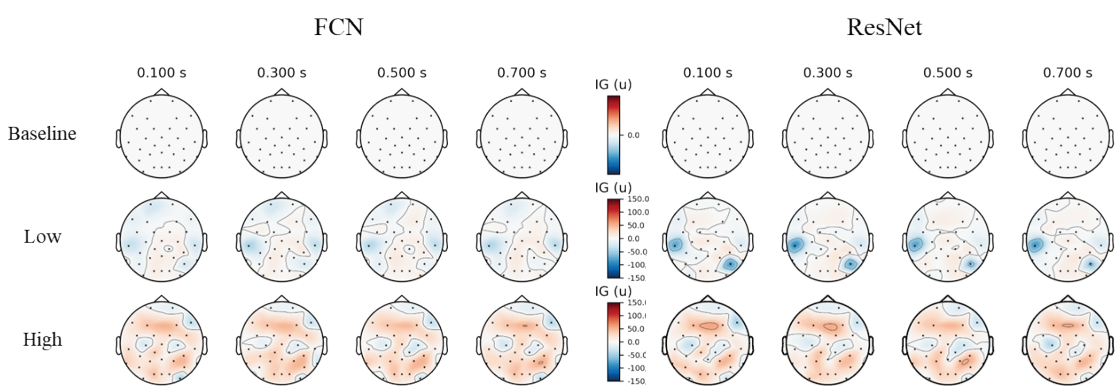
Integrated Gradients is a backpropagation-based method that assesses how a model's input features affect their predictions (Sundararajan et al., 2017). It computes the attribution of all input features to the output value via a forward and backward network traversal. Integrated Gradients have been judged to be an appropriate method for capturing global nonlinear effects and cross interactions between different features (Ancona et al., 2017), i.e., properties that match EEG recorded waveform characteristics such as the different range of oscillations, spatial distribution, and related synchronized activity. The technique is recommended as explanation methods in neurophysiological data by (Thomas et al., 2023). Moreover, Integrated Gradients respond to an important axiom related to neural network interpretability, i.e., "the notion of completeness" (Gilpin et al., 2018), which in this context, states that the sum of attributions should be equal to the original output. In brief, integrated gradients highlight areas of high network sensitivity where a change in signal input may significantly affect the output. We provide a note of caution that the results from this feature assessment are an oversimplification of what features a model has learned and do not represent the complexity of the models' behavior.

For FCN and ResNet we applied Integrated Gradients to our validation dataset to determine which signal features and sensors contributed to classification. The contributions were measured relative to a baseline, which offered no information to the models. To ensure the representativeness of our results, we selected to baseline using two factors following (Sundararajan et al., 2017). 1) we use a baseline that conveys a total absence of signal using an all-zero multidimensional vector, and 2) use samples of participants' EEG during resting phases with eyes-open recorded prior to all tasks. Research has shown that baseline selection can drastically change feature attributions and provide advice to select one depending on the domain and the task (Kindermans et al., 2019).

In order to make Integrated Gradients understandable, we computed the average marginal contribution of all EEG channels to the estimated high workload level per trial. Then, we produced a visualization of these gradients to illustrate positive contributions, where red shows a channel's features that positively influence the prediction toward the classified output. In blue, the negative contribution that influences the prediction away from the expected output. Figure 24 is the result of this procedure.

Figure 24

Topography of the integrated gradient values for each sensors overtime.



Note. It represents the grand average for each condition computed using the IG values averaged for all participants between the multiple folds. Note that these visualizations oversimplify what features a model has learned and do not represent the complexity of the models' behavior.

Previous research has linked mental workload to spectral markers of cortical activity in the parietal and frontal areas of the brain. Specifically, a decrease in the Alpha (8-12 Hz) frequency band in the parietal region, coupled with an increase in the Theta (4-8 Hz) and often Delta (1-4 Hz) bands in the frontal region, are commonly associated with an increase in mental workload (Antonenko et al., 2010; Borghini et al., 2014; Brouwer et al., 2012; Krause et al., 2000; Lei & Roetting, 2011; Roy et al., 2013).

Our results from the assessment using Integrated Gradients revealed that similar cortical areas contribute to the mental workload output classification. Specifically, the visualization of the values showed that both the frontal and parietal areas contribute towards the classification. Moreover, we observed that noise-prone electrodes, which may pick up non-brain correlated information such as the prefrontal lobe (e.g., Fp1 and Fp2), contributed little to the classification. These two observations provide us with confidence in the neurobiological plausibility of the features utilized by our models.

3.7.2 Model application

In order to apply the trained models in classification mode to the EEG signal data recorded during the flight training simulation and thus estimate the mental workload level for each participant, we segmented and epoched the data. These data were segmented by maneuver with an extracted 10% buffer before the start and after the end of each maneuver to avoid confounding factors that can affect brain activity, such as muscle movement potentially linked to a participant's posture or moving for comfort within the simulator between maneuvers. Segments were then partitioned into 3000ms epochs using a 1500ms overlapping window to correspond with the data epochs used to train the classifiers. As a result of the process, the low complexity maneuvers took between 26 and 60.67 epochs on average, while the medium complexity maneuvers took 61.11 to 84 epochs, and the high complexity maneuvers took 68.56 to 121.22 epochs (Table 27). It was expected that low-complexity maneuvers would be completed in a shorter time than the high-complexity maneuvers and thus result in a lower number of epochs per maneuver per participant. We then applied the trained models (i.e., FCN and

ResNet) to classify mental workload for each of the epochs. Before utilizing the estimations produced by our models, we conducted two validation procedures: inter-level agreement and intra-model agreement.

Table 27

Number of epochs per participant, block and maneuveurs

Block	Maneuver	1	2	3	4	5	6	7	8	9	Average
Low	B	26	19	35	32	36	25	24	15	22	26
	C	56	36	65	72	57	41	50	28	38	49.22
	D	58	37	70	32	81	76	36	31	50	52.33
	E	61	45	70	65	51	50	53	77	74	60.67
Moderate	F	96	3	115	154	68	58	65	42	75	75.11
	G	73	46	66	74	64	56	40	68	63	61.11
	H	39	69	121	66	72	85	52	45	63	68
	I	35	77	93	109	121	108	41	43	129	84
High	J	107	77	185	119	137	96	207	79	72	119.89
	K	100	135	178	131	94	147	85	97	124	121.22
	L	139	31	43	68	63	62	94	55	62	68.56

Note. Maneuvers starting from B to L was used. The training phase, maneuver A was discarded, as it serves as an adaptation period.

3.7.3 Inter-model agreement

Looking exclusively at the performance of individual models does not provide an indication of their generalizability. To assess the reliability of the models when applied to unseen data, we computed the agreement level between the models when classifying mental workload from the same EEG data epochs. We hypothesized that the two models with different designs might extract dissimilar features and leverage distinct information from the signal before computing the probabilistic classification outcome. However, if the two models agree on the estimation using similar features, it may indicate that discriminative information is present in the signal.

We use Cohen's Kappa to measure the inter-classifier agreement level on the same classification problem: $\kappa = \frac{p_o - p_e}{1 - p_e}$ were p_o represents the observed agreement and p_e the expected agreement by chance alone (Viera & Garrett, 2005). A kappa of 1 represents a perfect agreement, while a kappa of 0 is equivalent to chance. We also compute two correlation measures, the Concordance Correlation Coefficient (CCC) (Lawrence & Lin, 1989) and the Matthews Correlation Coefficient (MCC) (Matthews, 1975). The CCC represents the agreement between paired observations, enabling the comparison of two instruments intended to measure the same target. In our case, CCC measures the similarity of the probability in estimating the mental workload level for our two models, FCN and ResNet. The advantage of this measure is its applicability to continuous values. The MCC measures binary classification quality, considering true positives, false positives, true negatives, and false negatives. The MCC value ranges from -1, total disagreement, to 1, total agreement, with a value of 0 indicating random chance. In brief, MCC is a correlation measure for discrete observations, i.e., high or low workload.

Figure 25

Inter-Model Level Agreement

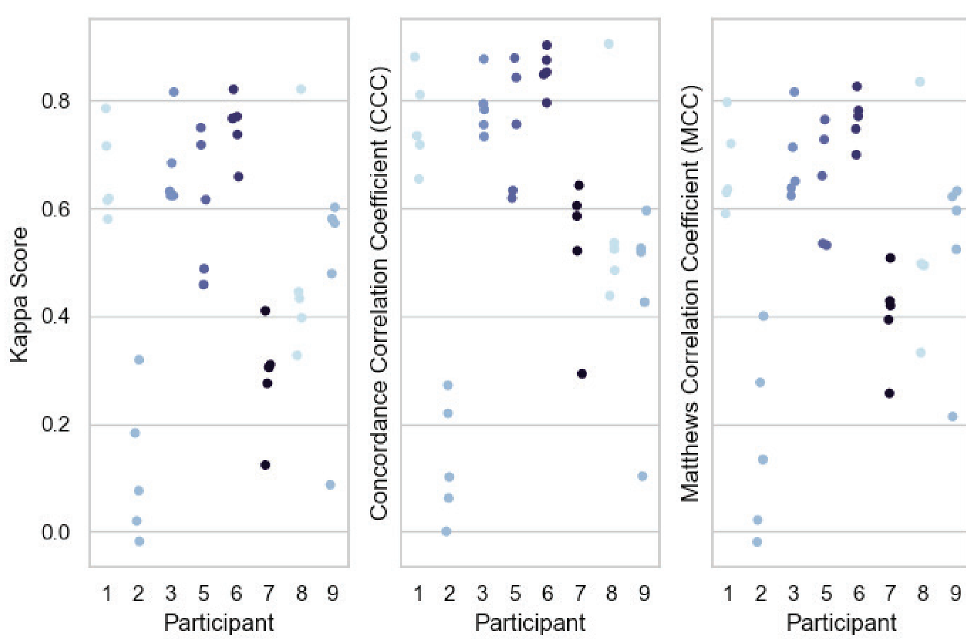


Figure 25 shows each participant's FCN and ResNet agreement scores on the same unseen data. The average scores for each metric were computed, along with their corresponding standard deviations, to measure the variability. The fourth participant was excluded from the analysis as the model consistently estimated a high workload level for all epochs. Upon careful consideration, this observation may be attributed to a potential drift in the distribution and magnitude of the EEG signal, likely caused by a decrease in signal quality between the synthetic the naturalistic task. Overall, the mean Kappa Score was 0.506 ± 0.216, indicating a moderate agreement between the predicted and actual classifications. The mean CCC was 0.603 ± 0.239, indicating moderate agreement between the models' predictions. For the MCC, the mean was calculated as 0.487, with a standard deviation of $\pm$ 0.257, suggesting a moderate level of prediction accuracy. We observed a decline in scores for the second participant, with an average Kappa Score of 0.116 ± 0.136, CCC of 0.131 ± 0.112, and MCC of 0.163 ± 0.175. A comparative assessment of the features using Integrated Gradient revealed that FCN and ResNet utilized similar scalp areas over time. However, the cause of this disagreement is still unclear. In this case, they could not reliably infer the level of mental workload over the novel EEG flight data for this participant.

3.7.4 Intra-model agreement

Random initialization and the choice of training folds can impact the performance of deep learning models when applied to EEG datasets. Inconsistent estimations within the same model raise concerns about the instrument's reliability. We controlled for these factors during the performance benchmark to address this issue by employing 5-fold training on subject-dependent datasets for each model and design choice.

However, we aimed to validate our models further and assess whether different initializations and training folds could affect the learned discriminative features and the prediction accuracy on unseen EEG data. Therefore, we employed the same approach as the intra-model agreement evaluation. For this assessment, we posed the question: "Will the models consistently estimate the level of mental workload if trained again?" We computed Cohen's Kappa, CCC, and MCC for FCN and ResNet models trained on the 5 folds and applied them to the unseen EEG data to answer this question. The scores were computed for each combination of the 5 models trained per participant.

For the FCN model, the mean Cohen's Kappa is 0.635 ± 0.198, the mean Concordance Correlation Coefficient (CCC) value is 0.756 ± 0.182, and the mean Mathew Correlation Coefficient (MCC) value is 0.590 ± 0.266. In the case of the ResNet model, the mean Cohen's Kappa is 0.590 ± 0.191, the mean CCC value is 0.654 ± 0.204, and the mean MCC value is 0.554 ± 0.249. Figures 26 and 27 illustrate the scores for each pair of participants and metrics. Overall, the scores indicate substantial to moderate agreement within each training of the models and folds.

Figure 26

Intra-Model Level Agreement for FCN

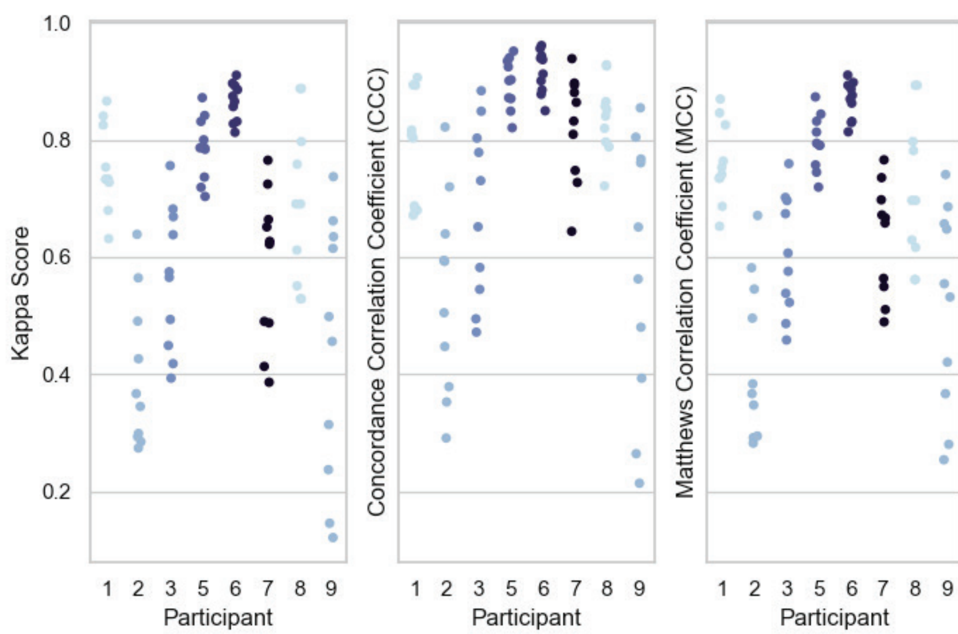
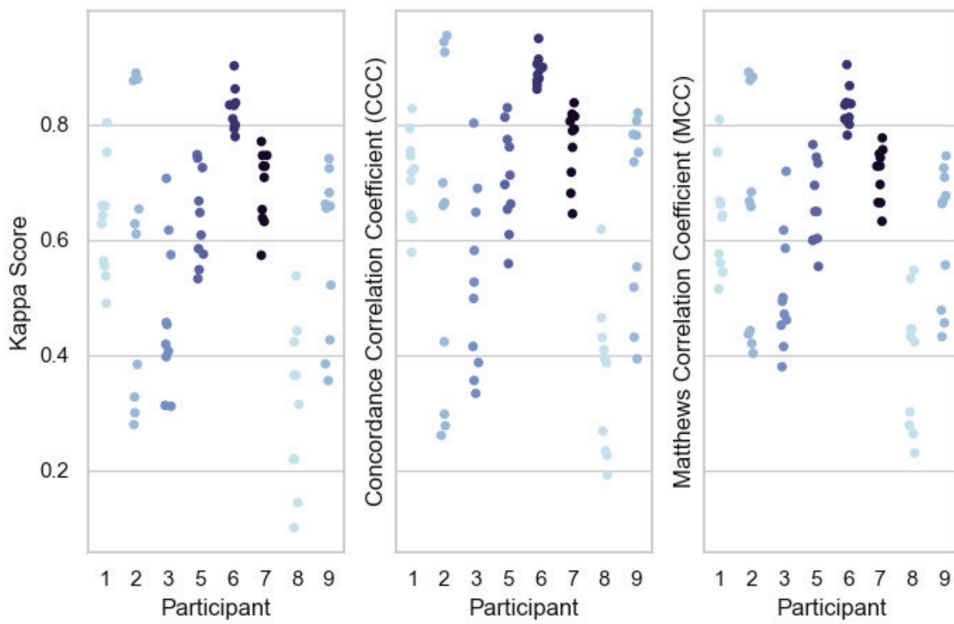


Figure 27

Intra-Model Level Agreement for ResNet



3.8 Results

We conducted a Linear Mixed Effects (LME) analysis to examine the relationship between flight maneuver complexity and participant performance during flight simulation training. This analysis was chosen since we had multiple measures per subject for each complexity level, violating the independence assumption of a linear regression model. We utilized the R programming language and the lme4 package (Bates et al., 2014) for performing the linear mixed effects analysis. Furthermore, we assessed the model's assumptions using the Performance package (Lüdecke et al., 2021).

We applied both models to the EEG data collected during the flight simulation task, and epochs were segmented according to the procedure outlined in the section on model application. The statistical analysis used the performance scores assessed by a subject matter expert as the dependent variable for each maneuver corresponding to the three complexity levels. Mean Estimated Mental Workload (MEMW) was used as a predictor, representing the average probability of estimated mental workload over all epochs of the maneuvers corresponding to a given complexity level. A value of 1 indicates full confidence in the model's estimation of high mental workload, while a value of 0 indicates full confidence in the model's estimation of low mental workload. Estimated Mental Workload Proportion (EMWP) was also used as a predictor, representing the proportion of epochs classified as high mental workload, with a value of 1 signifying all epochs classified as high, a value of 0 signifying all epochs classified as low. Table 28 and Figure 28 present the descriptive statistics of each of these variables per complexity level.

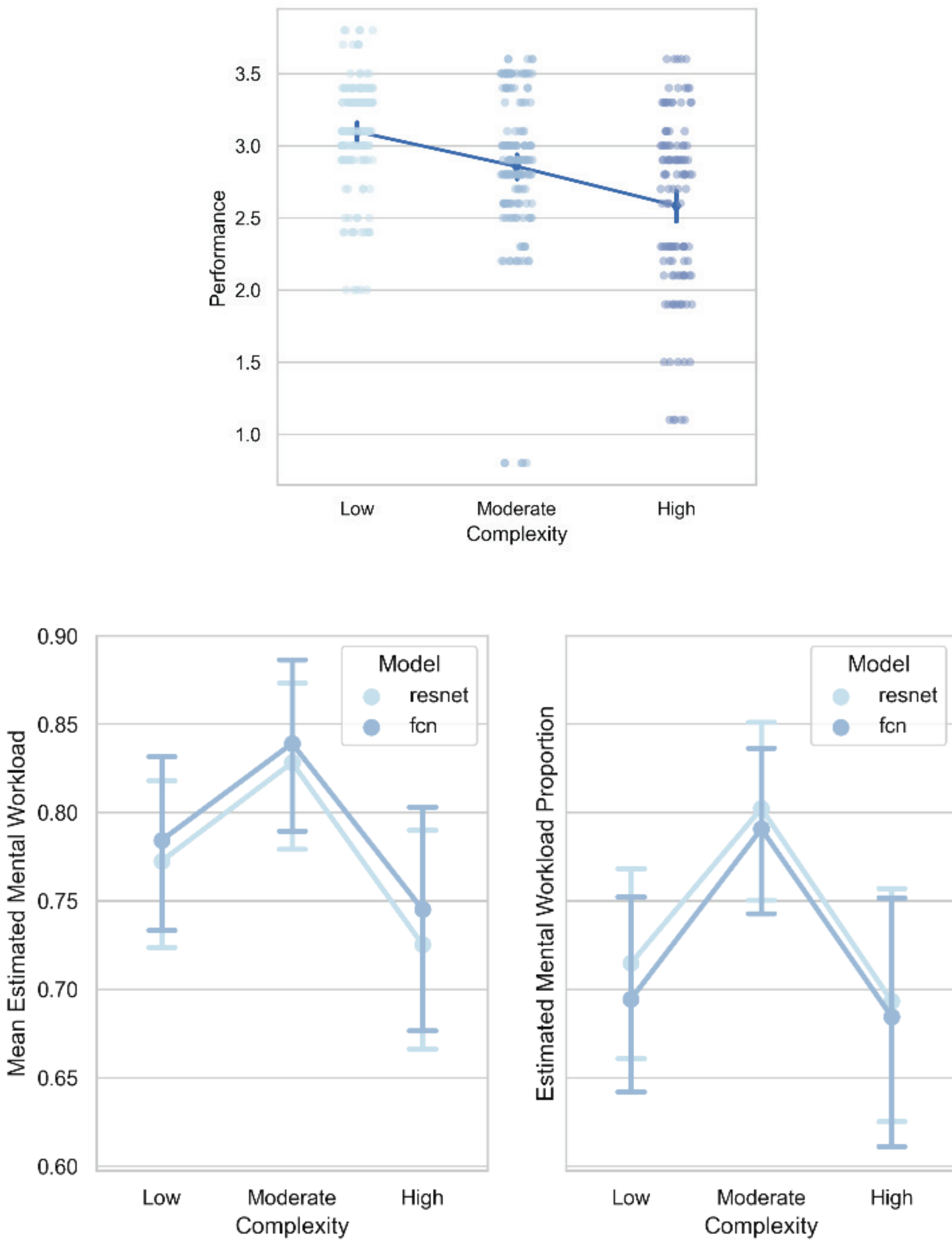
Table 28

Descriptive Statistics of the measures

Model	Measure	Complexity		
		Low	Moderate	High
	Performance	3.097 ± .379	2.856 ± .533	2.583 ± .612
FCN	MEMW	.784 ± .315	.839 ± .300	.745 ± .360
	EMWP	.695 ± .386	.791 ± .322	.685 ± .389
Resnet	MEMW	.773 ± .293	.828 ± .299	.725 ± .357
	EMWP	.715 ± .339	.802 ± .316	.693 ± .378

Figure 28

Visual representation of the performance, MEMW, and EMWP per complexity level for both FCN and ResNet models.



To apply the LME, and to determine the appropriate maximal random effects structure, we followed the procedure outlined by Barr et al. (2013). Given our dataset, which includes multiple performance observations per subject and complexity level, we considered the possibility that the performance may be influenced by the manipulated factor—the complexity of the maneuver—and that this effect might vary across subjects. Similar considerations are appropriate considering estimated mental workload. To account for these factors, we specified a random gaussian intercept and random slope for subjects, allowing them to be influenced by the complexity of the maneuver. The LME models were estimated using the restricted maximum likelihood (REML) method with the nloptwrap optimizer. We calculated 95% confidence intervals and p-values to assess statistical significance using a Wald t-distribution approximation. Additionally, residual plots were visually examined for deviations from homoscedasticity or normality, and no significant deviations were observed. These plots are provided in the appendix for reference. To present our findings, we computed five LME models. Model 1 examined the association between complexity and performance. Models 2 and 3 included the fixed effects of the mental workload estimations (EMWP and MEMW) obtained from the ResNet model. Similarly, Models 4 and 5 incorporated the same fixed effects using the FCN model.

The results indicate that maneuver complexity is significantly associated with performance, with higher complexity levels being negatively related to evaluated performance per maneuver. The effect of complexity was statistically significant and negative ($\beta = -0.26$, 95% CI [-0.41, -0.11], $t(82) = -3.40$, $p = 0.001$; Std. $\beta = -0.37$, 95% CI [-0.59, -0.15]).

Next, we included EMWP and MEMW as fixed effects from both models. For the ResNet model, no significant association was found between EMWP or MEMW and performance. The effect of EMWP was statistically nonsignificant and negative ($\beta = -0.10$, 95% CI [-0.51, 0.32], $t(81) = -0.47$, $p = 0.642$; Std. $\beta = -0.06$, 95% CI [-0.30, 0.19]). Similarly, the effect of MEMW was statistically nonsignificant and negative ($\beta = -0.08$, 95% CI [-0.52, 0.36], $t(81) = -0.36$, $p = 0.723$; Std. $\beta = -0.04$, 95% CI [-0.29, 0.20]). Comparable results were observed for the association between EMWP, MEMW, and performance. The effect of MEMW was statistically nonsignificant and positive ($\beta = 0.06$, 95% CI [-0.38, 0.49], $t(81) = 0.26$, $p = 0.796$; Std. $\beta = 0.04$, 95% CI [-0.24, 0.31]). Likewise, the effect of EMWP was statistically nonsignificant and positive ($\beta = 0.07$, 95% CI [-0.33, 0.48], $t(81) = 0.37$, $p = 0.716$; Std. $\beta = 0.05$, 95% CI [-0.22, 0.31]).

As post-hoc analysis, in order to examine the association between perceived MW, serving as our manipulation check, and the estimated mental workload, we conducted Pearson's correlation analyses as part of an exploratory investigation. For the ResNet model, a moderate positive correlation emerged between perceived MW and estimated mental workload for EMWP ($r = 0.24$, 95% CI [0.03, 0.42], $t(86) = 2.25$, $p = 0.027$), while a weak positive correlation was observed between perceived MW and estimated mental workload for MEMW ($r = 0.22$, 95% CI [0.01, 0.41], $t(86) = 2.07$, $p = 0.041$). The correlations obtained from the FCN model were consistent with these findings. Specifically, a moderate positive correlation was identified between perceived MW and estimated mental workload for EMWP ($r = 0.25$, 95% CI [0.04, 0.44], $t(86) = 2.39$, $p = 0.019$), and a weak positive correlation was observed between perceived MW and estimated mental workload for MEMW ($r = 0.25$, 95% CI [0.04, 0.43], $t(86) = 2.37$, $p = 0.020$).

3.9 Discussion

3.9.1 Empirical implications

This work reviewed end-to-end deep learning techniques for mental workload estimation and benchmarked several to determine the best-performing architectures for the task. The results show that three deep learning architectures (FCN, ResNet, EEGNet) perform better than the state-of-the-art baselines (i.e., COV+RMNM). Furthermore, we systematically evaluated a number of end-to-end deep learning models, design choices, and architectures for mental workload estimation directly from the EEG signals. The two top-performing models ResNet and FCN produced an average accuracy of .917 (± .074) and .933 (± .054), respectively, over the training data.

Our endeavour was motivated by our review of the research concerning the application of deep learning to mental workload estimation, which showed that the majority of studies trained models using handcrafted and domain-specific engineered feature data. These feature data such as Fast Fourier Transform (Casson, 2014; Jiao et al., 2018; Kuanar, Athitsos, Pradhan, Mishra, & Rao, 2018; Qiao & Bi, 2020; Shang et al., 2017; Tao et al., 2019; Yang et al., 2019), Short-Time Fourier Transform (Hefron, Borghetti, Kabban, et al., 2018; Wang et al., 2012), Wavelet Transform approaches (Qayyum, Faye, et al., 2018; Wu et al., 2019), Event-Related Desynchronization/Synchronisation (Saadati et al., 2020b), and Functional Connectivity with EEG graph representation (Wang et al., 2019) were used to transform the signal and input unaltered or as image representation to deep neural networks for mental workload (or similar) estimation tasks, producing fair accuracies. However, feature engineering generally requires domain-specific expertise, additional computing resources and time, and does not consider individual variance in neurophysiological responses (Donoghue et al., 2020; Donoghue et al., 2021). Furthermore, feature engineering does not leverage the strength of recent developments in end-to-end processes for discriminative models using time series data. Models that utilize end-to-end processes can learn hierarchical representations directly from raw signal data to provide a probabilistic inference of the class prediction without assuming a prior domain knowledge representation of the EEG signal. Our work contributes to the literature by providing evidence of the performance of a wide selection of unique deep learning architecture for mental workload estimation.

In terms of the results of the estimation of mental workload for the simulated flight task using the classification output from the FCN and ResNet models we trained using n-back task data, we found the relationship between complexity, mental workload and performance to be a u-shape relationships in line with previously reported research (Svensson et al., 1997). Firstly, we found that the simulator training scenarios, through the manipulation of flight maneuver complexity, could elicit a linear increase in perceived mental workload (manipulation check: RAW-TLX), building a solid foundation of validity for subsequent results. Secondly, we found that performance was negatively associated with the complexity of the task, however we find no relationships with the estimated mental workload.

Table 29

Research implications

Element	Implication
Empirical	ResNet and FCN end-to-end deep learning perform better than state-of-the-art baselines of the same nature
	Benchmark of end-to-end deep learning architectures shows varying performances on mental workload estimation showing the importance of testing several models
	We show convergence between features learned by the models and the neurophysiological plausibility
Methodological	We propose a methodological framework for end-to-end deep learning benchmarking
	End-to-end deep learning models allow the user of transfer learning techniques that can increase further generalizability, reduce training time, and enable test-retest capability.
Theoretical	End-to-end deep learning models can offer theoretical insights on the relationship between physical responses, cognitive mechanisms and even behaviors.
	End-to-end deep learning models could serve as a rich and continuous measurement instrument of constructs in theory testing

We observed that high-complexity maneuvers did not result in an increase in the number of estimations of high mental workload across the three flight task blocks. This finding contradicted the perceived mental workload assessed through the NASA Task Load Index (TLX) at the end of each difficulty block and its relationship with performance (Svensson et al., 1997). Contrary to our expectations, we discovered a reverse U-shaped relationship between the increase in complexity and the estimated mental workload. Specifically, we observed a non-significant increase in the estimation of high workload between low and moderate levels of complexity, followed by a decline at the high level of complexity. However, post-hoc analysis shows a significant positive correlation between estimated cognitive mental workload and perceived mental workload showing that the relationships might be more complex than expected.

Although seemingly counterintuitive, these results can be interpreted in tree possible ways. Firstly, it could indicate a learning effect within the flight simulator and towards the flight tasks. Participants might have experienced an increase in perceived mental workload, but this may not have been accompanied by significant psychophysiological reactivity. Secondly, the results could align with previous studies on mental workload, which propose the existence of an upper limit. According to this interpretation, the higher complexity flight maneuvers pushed participants beyond this upper limit, leading to a state of cognitive overload. In such a state, additional complexity does not elicit further psychophysiological reactivity (Jaeggi et al., 2007). A third alternative explanation is that the reliability and generalizability of the mental workload measurement can be improved. Consequently, further work should be conducted to improve the measurement techniques.

3.9.2 Methodological implications

Nonetheless, end-to-end deep learning analysis can offer a powerful complementary alternative to traditional approaches to the classification of mental states. However, the approach creates a new set of challenges that must be addressed. As shown in the neurophysiological plausibility assessment, when using deep learning to learn features and estimate mental state directly from the signal, a model will utilize any discriminant feature regardless of whether that feature is related to the target cognitive state for classification (Kohoutová et al., 2020). Predictive artifacts can confound and affect the performance and accuracy (either positively or negatively) of the model, depending on the task.

As part of our methodology for model training, we performed a model assessment based on Integrated Gradients (Sundararajan et al., 2017) to compute the attribution of channels to the models' outputs compared to a baseline (i.e., a total absence of signal using an all-zero multidimensional vector). We attempt to relate the findings of the neural networks to the past literature on mental workload assessment to identify the results' neurophysiologic plausibility. We found known brain-related patterns, such as frontal and central parietal activations. In this assessment, we found significant consonance with the neurophysiological literature.

Our findings demonstrate the relevance of our methodological framework, highlighting its relevance in practice. It is crucial to verify the plausibility of the features learned by the models prior to their application, ensuring that the end-to-end deep learning models effectively utilize physical brain data. By formalizing this method, we contribute to the enrichment of existing approaches, providing a systematic and rigorous framework for evaluating the appropriateness of the learned features during the benchmarking process.

3.9.3 Theoretical implications

Thus, this approach holds the potential to make significant contributions to theoretical advancement in two distinct ways. Firstly, an expanding body of evidence highlights the agreement between neural networks and the brain's physical responses (Doerig et al., 2023; Thomas et al., 2023). Exploring the impact of architectural design choices on model learning can deepen our comprehension of the brain, electroencephalography (EEG), and their manifestations across various datasets, tasks, and relevant neural mechanisms. Furthermore, employing interpretation techniques to assess the neurophysiological plausibility of the acquired features provides valuable insights into how the models utilize brain data to make estimations, elucidating the relationship between physical responses, cognitive mechanisms and even behaviors (Doerig et al., 2023). Models trained in this way can potentially be used as a form of fundamental discovery when coupled with traditional neuroscience analysis methods.

Secondly, the utilization of end-to-end deep learning models presents an opportunity to employ them as measurement instruments for testing theoretical models. This approach offers a flexible means to operationalize and measure constructs in a longitudinal manner. By utilizing the output of the machine learning model as a measure within a statistical model, the temporal nature of the phenomena is preserved. Consequently, this methodology enables the examination of theories that account for mental states change over time, facilitating complex study of cognition in HCI.

3.9.4 Limitations and further work

Several limitations are associated with our current implementation of end-to-end deep learning processes. The first limitation is our sample of novice pilots. Their performance and mental workload could be influenced by their limited experience and learning curve, which might be affected as the complexity of maneuvers increases.

Secondly, we use a binary estimation of mental workload. We faced a trade-off between an increase in the discriminative power of the models (less class granularity for better generalization) and an increase in the confidence limit of the classifier to perform better than random. The binary class problem was chosen to increase the models' discriminative capability over the currently available brain data. Increasing the sample size might allow for more granularity in mental workload classification.

Thirdly, we could apply additional design choices and hyperparameters to improve performance and accuracy. For example, we showed that the residual architecture (i.e., ResNet) increased the predictive accuracy through parameter tuning, which compares favorably with similar findings (Hefron, Borghetti, Schubert Kabban, et al., 2018). However, in our review of the literature, we did not uncover similar research that utilized this architecture for mental workload estimation based on EEG signal data. Moreover, while we do not leverage the commonly engineered features used for mental workload estimation in this work, we evaluated the model's learned features through feature attribution to assess their validity for a neurophysiological inference of mental workload.

Moreover, advances in network design with the potential to improve accuracy and performance are being reported constantly in the rapidly advancing field of deep learning that can improve performance, e.g., the type of pooling layers, regularization, and normalization techniques. Thirdly, architectural design choices can be selected for specific properties of EEG signal data. For example, Gao et al. (2019) built upon the spatial and temporal dependencies of EEG signal data over spatial and spectral dependencies (Qiao & Bi, 2020), and (Hefron et al., 2017) showed that taking into account temporal dependencies in EEG data increased performance of models. Finally, instead of representing these signal data properties via engineered features, model architectures can be built to generate them, e.g., building a graph network for neural connectivity, using a recurrent neural network coupled with a convolutional neural network for temporal dependencies and features learning.

Thus far, we have also not attempted to train a fully generalizable subject-independent or time-independent model. Generalization across several days or the “test-retest” reliability of models remains an important challenge for mental state estimation using psychophysiological data (Wilson et al., 2010). According to Christensen et al. (2012) a model's capability to discriminate between different mental states decreases over time due to physiological factors. The non-stationarity of the target signal impedes the training and application of models at different points in time, even within-subject. Moreover, the heterogeneous nature of physiological responses across individuals hinders the creation of generalized deep-learning models and features. To tackle these challenges, the results from models benchmarked in this study support the use of transfer learning. Transfer learning is a technique that aims to train an original network on a first dataset and task and then transfer the learned features to a new network to be trained on a new dataset and task (Yosinski et al., 2014). Conditional to the first model to learn generalizable features, this practice can lead to better consistency in the performance of deep neural networks for training and testing within-subject across days or between-subjects when datasets of labeled data are limited.

During our review of deep learning techniques for mental workload, we identified one study utilizing transfer learning for image classification using discrete wavelet transform to discriminate during resting and learning phases (Qayyum, Faye, et al., 2018). For tasks other than mental workload estimation, transfer learning has shown promising results when applied to EEG signal data. Lin and Jung (2017) proposed a transfer learning pipeline for emotion recognition that assesses the comparability of brain response prior to tuning the inter-subject train model. Fahimi et al. (2019) applied transfer learning based on the training of a deep CNN across-subject for attention discrimination. Similar approaches have been used for drowsiness detection (Wei et al., 2018).

In our context, the synthetic n-back task and the recorded EEG signal data offer the opportunity to fine-tune a pre-trained model for mental workload estimation before its application to the ecologically valid simulated flight data affording faster training with less labeled EEG signal data windows. Our architectures allow for a cross-subject model that can be trained, then tuned and applied to the current subject. Finally, further work can focus on developing domain adaptation, a special form of transfer learning technique, to enhance classifier performance and generalizability when tasks are similar (Lotte et al., 2018).

3.10 Conclusion

In conclusion, this study demonstrates the feasibility of employing an end-to-end deep learning approach for classifying mental workload based on EEG signal data. The results show exciting results in estimating mental workload in naturalistic tasks and provide a path for further research in creating a reliable instrument. We argue that this approach holds significant potential as a powerful tool for the direct discrimination of mental states from raw data. However, it also presents inherent challenges that need to be addressed, such as conducting thorough analyses of the features and ensuring their neurophysiological plausibility. A challenge that might be exacerbated when the technique is applied to naturalistic environments due to even more noisy environments and potential body movements. Nonetheless, we believe that further research in estimating mental workload in naturalistic HCI tasks can lead to valuable scientific advancements. Building upon this research, evaluating EEG-specific design choices and incorporating transfer learning techniques for task-adaptive models can further enhance the performance and generalization of the estimation techniques for naturalistic tasks.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., & Devin, M. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems.

Almogbel, M. A., Dang, A. H., & Kameyama, W. (2018). EEG-signals based cognitive workload detection of vehicle driver using deep learning. 2018 20th *International Conference on Advanced Communication Technology* (ICACT),

Almogbel, M. A., Dang, A. H., & Kameyama, W. (2019). Cognitive Workload Detection from Raw EEG-Signals of Vehicle Driver using Deep Learning. In *2019 21st International Conference on Advanced Communication Technology* (pp. 1167-1172).

Almogbel, M. A., Dang, A. H., Kameyama, W., & Ieee. (2018). EEG-Signals Based Cognitive Workload Detection of Vehicle Driver using Deep Learning. In *2018 20th International Conference on Advanced Communication Technology* (pp. 256-259).

Almogbel, M. A., Dang, A. H., Kameyama, W., & Ieee. (2019). Cognitive Workload Detection from Raw EEG-Signals of Vehicle Driver using Deep Learning. In *2019 21st International Conference on Advanced Communication Technology* (pp. 1167-1172).

AlZoubi, O., Calvo, R. A., & Stevens, R. H. (2009). Classification of EEG for affect recognition: an adaptive approach. Australasian Joint Conference on Artificial Intelligence,

Ancona, M., Ceolini, E., Öztireli, C., & Gross, M. (2017). Towards better understanding of gradient-based attribution methods for deep neural networks. *arXiv preprint arXiv:1711.06104*.

Antonenko, P., Paas, F., Grabner, R., & Van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational psychology review*, *22*(4), 425-438.

Appriou, A., Cichocki, A., & Lotte, F. (2018). Towards robust neuroadaptive HCI: exploring modern machine learning methods to estimate mental workload from EEG signals. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems,

Ayaz, H., Çakir, M. P., İzzetoğlu, K., Curtin, A., Shewokis, P. A., Bunce, S. C., & Onaral, B. (2012). Monitoring expertise development during simulated UAV piloting tasks using optical brain imaging. 2012 IEEE Aerospace Conference,

Baldwin, C. L., & Penaranda, B. N. (2012). Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. *Neuroimage*, *59*(1), 48-56. https://doi.org/10.1016/j.neuroimage.2011.07.047

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Benavoli, A., Corani, G., & Mangili, F. (2016). Should we really use post-hoc tests based on mean-ranks? *The journal of machine learning research*, *17*(1), 152-161.
Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.
Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, *78*(5), B231-B244.
Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of ERP components—a tutorial. *Neuroimage*, *56*(2), 814-825.
Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, *44*, 58-75.
Brouwer, A.-M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering*, *9*(4), 045008.
Burke, J. L., Murphy, R. R., Rogers, E., Lumelsky, V. J., & Scholtz, J. (2004). Final report for the DARPA/NSF interdisciplinary study on human-robot interaction. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *34*(2), 103-112.
Casson, A. J. (2014). Artificial Neural Network classification of operator workload with an assessment of time variation and noise-enhancement to increase performance. *Frontiers in neuroscience*, *8*, Article 372. https://doi.org/10.3389/fnins.2014.00372
Chen, X., Li, C., Liu, A., McKeown, M. J., Qian, R., & Wang, Z. J. (2022). Toward open-world electroencephalogram decoding via deep learning: A comprehensive survey. *IEEE Signal Processing Magazine*, *39*(2), 117-134.
Chollet, F. (2015). keras. In.
Christensen, J. C., Estepp, J. R., Wilson, G. F., & Russell, C. A. (2012). The effects of day-to-day variability of physiological data on operator functional state classification. *Neuroimage*, *59*(1), 57-63.
Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering*, *16*(3), Article 031001. https://doi.org/10.1088/1741-2552/ab0ab5
Cui, X., Zhang, J., & Wang, R. (2016). Identification of Mental Workload Using Imbalanced EEG Data and.DySMOTE-based Neural Network Approach. *Ifac Papersonline*, *49*(19), 567-572. https://doi.org/10.1016/j.ifacol.2016.10.627
Dahl, G. E., Sainath, T. N., & Hinton, G. E. (2013). Improving deep neural networks for LVCSR using rectified linear units and dropout. 2013 IEEE international conference on acoustics, speech and signal processing,

Debie, E., Rojas, R. F., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., Garratt, M., & Abbass, H. A. (2019). Multimodal fusion for objective assessment of cognitive workload: a review. *IEEE transactions on cybernetics*.

Doerig, A., Sommers, R. P., Seeliger, K., Richards, B., Ismael, J., Lindsay, G. W., Kording, K. P., Konkle, T., Van Gerven, M. A., & Kriegeskorte, N. (2023). The neuroconnectionist research programme. *Nature reviews neuroscience*, 1-20.

Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A. H., Wallis, J. D., & Knight, R. T. (2020). Parameterizing neural power spectra into periodic and aperiodic components. *Nature neuroscience*, *23*(12), 1655-1665.

Donoghue, T., Schaworonkow, N., & Voytek, B. (2021). Methodological considerations for studying neural oscillations. *European Journal of Neuroscience*.

Dozat, T. (2016). Incorporating nesterov momentum into adam.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, *12*(Jul), 2121-2159.

Durantin, G., Gagnon, J.-F., Tremblay, S., & Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural Brain Research*, *259*, 16-23.

Fahimi, F., Zhang, Z., Goh, W. B., Lee, T.-S., Ang, K. K., & Guan, C. (2019). Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. *Journal of neural engineering*, *16*(2), 026007.

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, *33*(4), 917-963.

Fukuda, N., Nambu, I., & Wada, Y. (2019). Classification of Movement Direction From Electroencephalogram During Working Memory Time. *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2019*, 3107-3110. https://doi.org/10.1109/embc.2019.8857943

Gao, Z., Wang, X., Yang, Y., Mu, C., Cai, Q., Dang, W., & Zuo, S. (2019). EEG-Based Spatio-Temporal Convolutional Neural Network for Driver Fatigue Evaluation. *Ieee Transactions on Neural Networks and Learning Systems*, *30*(9), 2755-2763. https://doi.org/10.1109/tnnls.2018.2886414

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA),

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. Proceedings of the fourteenth international conference on artificial intelligence and statistics,

Grimes, D., Tan, D. S., Hudson, S. E., Shenoy, P., & Rao, R. P. (2008). Feasibility and pragmatics of classifying working memory load with an electroencephalograph. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. Proceedings of the human factors and ergonomics society annual meeting,

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). Elsevier.

Hefron, R., Borghetti, B., Kabban, C. S., Christensen, J., & Estepp, J. (2018). Cross-Participant EEG-Based Assessment of Cognitive Workload Using Multi-Path Convolutional Recurrent Neural Networks. *Sensors*, *18*(5), Article 1339. https://doi.org/10.3390/s18051339

Hefron, R., Borghetti, B., Schubert Kabban, C., Christensen, J., & Estepp, J. (2018). Cross-Participant EEG-Based Assessment of Cognitive Workload Using Multi-Path Convolutional Recurrent Neural Networks. *Sensors*, *18*(5), 1339.

Hefron, R. G., Borghetti, B. J., Christensen, J. C., & Kabban, C. M. S. (2017). Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation. *Pattern Recognition Letters*, *94*, 96-104.

Hua, C., Wang, H., Chen, J., Zhang, T., Wang, Q., & Chang, W. (2019). Novel functional brain network methods based on CNN with an application in proficiency evaluation. *Neurocomputing*, *359*, 153-162. https://doi.org/10.1016/j.neucom.2019.05.088

Jaeggi, S. M., Buschkuehl, M., Etienne, A., Ozdoba, C., Perrig, W. J., & Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(2), 75-89.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394-412.

Jiao, Z., Gao, X., Wang, Y., Li, J., & Xu, H. (2018). Deep Convolutional Neural Networks for mental load classification based on EEG data. *Pattern Recognition*, *76*, 582-595. https://doi.org/10.1016/j.patcog.2017.12.002

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163-178.

Kim, J., Kim, M.-K., Wallraven, C., Kim, S.-P., & Ieee. (2014). *Across-subject estimation of 3-back task performance using EEG signals*.

Kindermans, P.-J., Hooker, S., Adebayo, J., Alber, M., Schütt, K. T., Dähne, S., Erhan, D., & Kim, B. (2019). The (un) reliability of saliency methods. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (pp. 267-280). Springer.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Klosterman, S. L., & Eepp, J. R. (2019). Investigating Ensemble Learning and Classifier Generalization in a Hybrid, Passive Brain-Computer Interface for Assessing Cognitive Workload. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2019*, 3543-3546. https://doi.org/10.1109/embc.2019.8857882

Kohoutová, L., Heo, J., Cha, S., Lee, S., Moon, T., Wager, T. D., & Woo, C.-W. (2020). Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nature protocols*, *15*(4), 1399-1435.
Krause, J. B., Taylor, J. G., Schmidt, D., Hautzel, H., Mottaghy, F. M., & Muller-Gartner, H. W. (2000). Imaging and neural modelling in episodic and working memory processes. *neural networks*, *13*(8-9), 847-859. https://doi.org/10.1016/s0893-6080(00)00068-x
Kuanar, S., Athitsos, V., Pradhan, N., Mishra, A., & Rao, K. (2018). Cognitive Analysis of Working Memory Load from EEG, by a Deep Recurrent Neural Network. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),
Kuanar, S., Athitsos, V., Pradhan, N., Mishra, A., & Rao, K. R. (2018, April). Cognitive analysis of working memory load from EEG, by a deep recurrent neural network. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2576-2580). IEEE.
Kwak, Y., Song, W.-J., Kim, S.-E., & Ieee. (2019). Classification of Working Memory Performance from EEG with Deep Artificial Neural Networks. In *2019 7th International Winter Conference on Brain-Computer Interface* (pp. 149-151).
Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of neural engineering*, *15*(5), 056013.
Lawrence, I., & Lin, K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 255-268.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.
Lei, S., & Roetting, M. (2011). Influence of task combination on EEG spectrum modulation for driver workload estimation. *Human Factors*, *53*(2), 168-179.
Lin, Y.-P., & Jung, T.-P. (2017). Improving EEG-based emotion classification using conditional transfer learning. *Frontiers in Human Neuroscience*, *11*, 334.
Liu, Y., & Liu, Q. (2017). Convolutional Neural Networks with Large-Margin Softmax Loss Function for Cognitive Load Recognition. In T. Liu & Q. Zhao (Eds.), *Proceedings of the 36th Chinese Control Conference* (pp. 4045-4049).
Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving. *Frontiers in Human Neuroscience*, *13*.
Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, *15*(3), 031005.
Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60).
Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, *405*(2), 442-451.

Nguyen, A., Yosinski, J., & Clune, J. (2015). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. Proceedings of the IEEE conference on computer vision and pattern recognition,

Nweke, H. F., Teh, Y. W., Al-Garadi, M. A., & Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*, *105*, 233-261.

Plis, S. M., Hjelm, D. R., Salakhutdinov, R., Allen, E. A., Bockholt, H. J., Long, J. D., Johnson, H. J., Paulsen, J. S., Turner, J. A., & Calhoun, V. D. (2014). Deep learning for neuroimaging: a validation study. *Frontiers in neuroscience*, *8*, 229.

Qayyum, A., Faye, I., Malik, A. S., Mazher, M., & Ieee. (2018). Assessment of Cognitive Load using Multimedia Learning and Resting States with Deep Learning Perspective. In *2018 Ieee-Embs Conference on Biomedical Engineering and Sciences* (pp. 600-605).

Qayyum, A., Khan, M. K. A. A., Mazher, M., Suresh, M., & Ieee. (2018). Classification of EEG Learning and Resting States using 1D-Convolutional Neural Network for Cognitive Load Assesment. In *2018 Ieee Student Conference on Research and Development*.

Qiao, W., & Bi, X. (2020). Ternary-task convolutional bidirectional neural turing machine for assessment of EEG-based cognitive workload. *Biomedical Signal Processing and Control*, *57*, Article 101745. https://doi.org/10.1016/j.bspc.2019.101745

Riedl, R., & Léger, P.-M. (2016). Fundamentals of NeuroIS. *Studies in Neuroscience, Psychology and Behavioral Economics. Springer, Berlin, Heidelberg*.

Rivet, B., Souloumiac, A., Attina, V., & Gibert, G. (2009). xDAWN algorithm to enhance evoked potentials: application to brain–computer interface. *IEEE Transactions on biomedical engineering*, *56*(8), 2035-2043.

Rojas, R. F., Debie, E., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., Garratt, M., & Abbass, H. (2020). Electroencephalographic Workload Indicators During Teleoperation of an Unmanned Aerial Vehicle Shepherding a Swarm of Unmanned Ground Vehicles in Contested Environments. *Frontiers in neuroscience*.

Roy, R. N., Bonnet, S., Charbonnier, S., & Campagne, A. (2013). Mental fatigue and working memory load estimation: interaction and implications for EEG-based passive BCI. 2013 35th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC),

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, *16*(5), Article 051001. https://doi.org/10.1088/1741-2552/ab260c

Roy, Y., Hubert, B., Isabela, A., Alexandre, G., & Jocelyn, F. (2019). Deep learning-based electroencephalography analysis: a systematic review. *arXiv preprint arXiv:1901.05498*.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Saadati, M., Nelson, J., & Ayaz, H. (2020a). Convolutional Neural Network for Hybrid fNIRS-EEG Mental Workload Classification. In H. Ayaz (Ed.), *Advances in Neuroergonomics and Cognitive Engineering* (Vol. 953, pp. 221-232). https://doi.org/10.1007/978-3-030-20473-0_22

Saadati, M., Nelson, J., & Ayaz, H. (2020b). Multimodal fNIRS-EEG Classification Using Deep Learning Algorithms for Brain-Computer Interfaces Purposes. In H. Ayaz (Ed.), *Advances in Neuroergonomics and Cognitive Engineering* (Vol. 953, pp. 209-220). https://doi.org/10.1007/978-3-030-20473-0_21

Samima, S., & Sarma, M. (2019). EEG-Based Mental Workload Estimation. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2019*, 5605-5608. https://doi.org/10.1109/embc.2019.8857164

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping*, *38*(11), 5391-5420.

Shang, J., Zhang, W., Xiong, J., & Liu, Q. (2017). Cognitive Load Recognition Using Multi-channel Complex Network Method. In F. Cong, A. Leung, & Q. Wei (Eds.), *Advances in Neural Networks, Pt I* (Vol. 10261, pp. 466-474). https://doi.org/10.1007/978-3-319-59072-1_55

Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., & Mehler, B. (2014). Classifying driver workload using physiological and driving performance data: two field studies. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929-1958.

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. Proceedings of the 34th International Conference on Machine Learning-Volume 70,

Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013). On the importance of initialization and momentum in deep learning. International conference on machine learning,

Svensson, E., Angelborg-Thanderez, M., Sjöberg, L., & Olsson, S. (1997). Information complexity-mental workload and performance in combat aircraft. *Ergonomics*, *40*(3), 362-380.

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational psychology review*, 1-32.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*(3), 251-296.

Tao, J., Yin, Z., Liu, L., Tian, Y., Sun, Z., & Zhang, J. (2019). Individual-Specific Classification of Mental Workload Levels Via an Ensemble Heterogeneous Extreme Learning Machine for EEG Modeling. *Symmetry-Basel*, *11*(7), Article 944. https://doi.org/10.3390/sym11070944

Thomas, A. W., Ré, C., & Poldrack, R. A. (2023). Benchmarking explanation methods for mental state decoding with deep learning models. *Neuroimage*, *273*, 120109.
Thomas, L. C., & Wickens, C. D. (2001). Visual displays and cognitive tunneling: Frames of reference effects on spatial judgments and change detection. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,
Veltman, J., & Gaillard, A. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, *41*(5), 656-669.
Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam med*, *37*(5), 360-363.
vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a neurois research agenda with four areas of societal contributions. *European Journal of Information Systems*, *29*(1), 9-24.
Wang, M., El-Fiqi, H., Hu, J., & Abbass, H. A. (2019). Convolutional Neural Networks Using Dynamic Functional Connectivity for EEG-Based Person Identification in Diverse Human States. *Ieee Transactions on Information Forensics and Security*, *14*(12), 3259-3272. https://doi.org/10.1109/tifs.2019.2916403
Wang, Z., Hope, R. M., Wang, Z., Ji, Q., & Gray, W. D. (2012). Cross-subject workload classification with a hierarchical Bayes model. *Neuroimage*, *59*(1), 64-69. https://doi.org/10.1016/j.neuroimage.2011.07.094
Wang, Z., Yan, W., & Oates, T. (2017). Time series classification from scratch with deep neural networks: A strong baseline. 2017 international joint conference on neural networks (IJCNN),
Wei, C.-S., Lin, Y.-P., Wang, Y.-T., Lin, C.-T., & Jung, T.-P. (2018). A subject-transfer framework for obviating inter-and intra-subject variability in EEG-based drowsiness detection. *Neuroimage*, *174*, 407-419.
Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science*, *11*(4), 128-133.
Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors*, *45*(3), 360-380.
Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, *12*(1), 3-18.
Wilson, G. F., Russell, C., Monnin, J., Estepp, J., & Christensen, J. (2010). How does day-to-day variability in psychophysiological data affect classifier accuracy? Proceedings of the Human Factors and Ergonomics Society Annual Meeting,
Wu, E. Q., Peng, X. Y., Zhang, C. Z., Lin, J. X., & Sheng, R. S. F. (2019). Pilots' Fatigue Status Recognition Using Deep Contractive Autoencoder Network. *Ieee Transactions on Instrumentation and Measurement*, *68*(10), 3907-3919. https://doi.org/10.1109/tim.2018.2885608
Yang, S., Yin, Z., Wang, Y., Zhang, W., Wang, Y., & Zhang, J. (2019). Assessing cognitive mental workload via EEG signals and an ensemble deep learning classifier based on denoising autoencoders. *Computers in Biology and Medicine*, *109*, 159-170. https://doi.org/10.1016/j.compbiomed.2019.04.034

Yin, Z., & Zhang, J. (2016). Recognition of Cognitive Task Load Levels Using Single Channel EEG and Stacked Denoising Autoencoder. In J. Chen & Q. Zhao (Eds.), *Proceedings of the 35th Chinese Control Conference 2016* (pp. 3907-3912).

Yin, Z., & Zhang, J. (2017a). Cross-session classification of mental workload levels using EEG and an adaptive deep learning model. *Biomedical Signal Processing and Control*, *33*, 30-47.

Yin, Z., & Zhang, J. (2017b). Cross-subject recognition of operator functional states via EEG and switching deep belief networks with adaptive weights. *Neurocomputing*, *260*, 349-366. https://doi.org/10.1016/j.neucom.2017.05.002

Yin, Z., & Zhang, J. (2018). Task-generic mental fatigue recognition based on neurophysiological signals and dynamical deep extreme learning machine. *Neurocomputing*, *283*, 266-281. https://doi.org/10.1016/j.neucom.2017.12.062

Yin, Z., Zhao, M., Zhang, W., Wang, Y., Wang, Y., & Zhang, J. (2019). Physiological-signal-based mental workload estimation via transfer dynamical autoencoders in a deep learning framework. *Neurocomputing*, *347*, 212-229. https://doi.org/10.1016/j.neucom.2019.02.061

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? Advances in neural information processing systems,

Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.

Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2016). Exploiting multi-channels deep convolutional neural networks for multivariate time series classification. *Frontiers of Computer Science*, *10*(1), 96-112.

Zhou, Y., Huang, S., Xu, Z., Wang, P., Wu, X., & Zhang, D. (2021). Cognitive workload recognition using EEG signals and machine learning: A review. *Ieee Transactions on Cognitive and Developmental Systems*.

Chapter 4
Essay #3 – Oriented-Attention Measurement in Multisensory Human-Computer Interaction using Electroencephalography

Abstract

This paper presents a methodological approach to infer attentional orienting neurophysiological responses during human-computer interaction (HCI) using a multisensory perturbation technique. We show that perturbation paradigms can elicit an attentional-orienting response (ERP) that is robust enough to be measured and sensible to intertwined attentional factors such as top-down forces. We also show that this response is influenced by the naturalistic environment while still being sensitive to attentional demand. We demonstrate that our approach is robust to complex multisensory environments while retaining sensitivity to task properties and attentional demands. This paper shows the utility of neuroscience methods and mental state inference to evaluate technological artifact design choices.

4.1 Introduction

Human-computer interactions are inherently multisensory and exert pressures on users' attentional mechanisms. For instance, the sound and visual pop-up of an email on the screen, the vibration and sound of a phone ringing on a table, or colleagues conversing in the background. In this context, users employ their ability to enhance or inhibit attentional resources to optimize the brain's processing information emerging from our environment (Stein et al., 2014). The relevant cognitive mechanisms are automatically and unconsciously utilized by technology users in their daily lives at work, on computers, during hedonic technological activities, and in various other situations. Users are constantly inundated with induced multisensory cues affecting numerous sensory modalities (e.g., auditory, visual, somatosensory) originating from dynamic, complex, and ever-changing work environments, technology, or tasks. Multisensory integration helps humans navigate these contexts by solving the binding and causal inference required to function effectively, determining which sensory cue originates from which common event (Noppeney, 2021).

Multisensory cues can take many forms and profoundly influence our experience with technology. For example, spatiotemporally aligned multisensory stimuli have a higher chance of being prioritized for further processing and tend to capture humans' attention (Driver, 1996; Talsma et al., 2010; Van der Burg et al., 2008, 2009) and increase reaction times (Diederich & Colonius, 2004; Hagmann & Russo, 2016). Thus, it can enhance our detection performance as much as they can distract us from the task. It has direct application in HCI design; for example (Ho et al., 2007) compared the design of unimodal auditory, unimodal vibrotactile, and audio-tactile collision warning while driving. The results showed lower reaction times in drivers braking responses following the multimodal warning signals compared to the unimodal ones. These cues can also have detrimental effects on users' performance; they can be punctual salient events in the multisensory environment and can be perceived as interruptions. For example, (Addas & Pinsonneault, 2015) proposed a taxonomy of interruption showing that intrusion, task-irrelevant interruption (pop-up, electronic message, system message) can be detrimental to performance due to time consumption, attentional switching cost, or increased error rates. They can also be interventions, task-relevant interruption, and be beneficial to performance. Such technology-mediated interruptions have consequences on performance (Addas & Pinsonneault, 2018; Chen & Karahanna, 2018; Galluch et al., 2015). Future technologies might also combine sensory modalities to enhance interaction through the fusion of digital and physical cues such as volumetric displays to combine vision and touch, focused ultrasound to stimulate senses of touch and hearing, and olfactory technologies (Cornelio et al., 2021).

The careful consideration of the mechanisms of attention and multisensory integration is crucial for understanding the impact of HCI on users. In naturalistic tasks and environments, technology users are subjected to the salience of ongoing concurrent events while exerting executive control to maintain attentional resources on the task at hand (Matusz et al., 2019). However, the cognitive mechanisms involved in multisensory integration and its interplay with attention in real-world human-computer interaction remains little explored. Multisensory integration is typically studied in controlled laboratory settings, but hypotheses tested in such environments poorly generalize to naturalistic tasks and settings (Matusz et al., 2019). Gaining insights into these processes and developing methods to study them in HCI tasks could significantly enhance our understanding of technology's implications on users at the cognitive mechanism level for NeuroIS. Furthermore, it could offer innovative techniques for evaluating artifacts in design science research.

One methodology to explore the mechanisms linked to the multisensory integration of technological/environmental cues and attention is to measure neurophysiological response via an electroencephalogram (EEG) (Müller-Putz et al., 2015; vom Brocke et al., 2020). EEG is the dominant method used in NeuroIS research due to a variety of factors including its high spatial resolution, temporal resolution, cost-effectiveness, and availability of a comprehensive knowledge base (Müller-Putz et al., 2015; Riedl et al., 2014; Riedl et al., 2020). Unfortunately, the naturalness of HCI tasks creates challenges with EEG in generalizing measurements and hypotheses applied and tested in laboratory paradigms that maximize internal validity (Matusz et al., 2019). Cognitive mechanisms are studied with experimental stimuli that manipulate well-defined brain functions and attempt to control for confounding brain processes (e.g., motor control, motion, sensory stimulation). However, the findings supported by these paradigms do not generalize well to more complex and natural tasks or show different neurophysiological responses (Felsen & Dan, 2005; Matusz et al., 2019; Northoff, 2018). This limitation is not restricted to brain states but also behaviors (Ladouce et al., 2016). That is, artificial stimuli produce different behaviors in controlled contexts than stimuli presented in situ.

Thus, the challenge resides in finding neurophysiological responses linked to relevant cognitive mechanisms which are robust within naturalistic environments, human-computer interaction, and tasks. In this research, we build on the neurophysiological response of attentional orienting to multimodal cues. Evidence indicates it is robust to complex environments and paradigms (Burns & Fairclough, 2015; Ladouce et al., 2019; Zink et al., 2016) and is sensitive to attentional mechanisms (Macaluso et al., 2016; Talsma et al., 2010; Tang et al., 2016).

Attentional orienting corresponds to an unintentional shift of attention toward a sensory event (Schröger & Wolff, 1998). Using task-irrelevant distracting stimuli, we can study the difference in processing resources allocated in the orientation of attention toward a perturbation. The technique involved using a perturbation of sensory inputs to trigger a reorientation of attention toward it. This attentional drift generates a neurophysiological response measurable with EEG under the form of an Event-Related Potential (ERP). This technique has been applied in naturalistic environments and tasks. Ladouce et al. (2019) revealed that neural correlates of attention toward an auditory stimulation are reduced in real-world and natural behavior like walking in a dynamic environment. Attentional processes are measurable with ERPs to auditory perturbation and robust to real-world settings. Burns and Fairclough (2015) used auditory perturbation while subjects were playing games and showed that the neurophysiological response was sensible to the game difficulty. With a similar paradigm, Zink et al. (2016) showed that ERPs were influenced by the increased cognitive load of the real-world and the active motion induced by the task, i.e., pedaling. Evidence shows that neurophysiological responses triggered by a temporary reorientation of attention toward a perturbation are robust but influenced by real-world environments and tasks.

Therefore, developing such measurements necessitates consideration of the task and the environment in which it occurs. Consequently, conceptualizing the interaction between the user, the brain, and the environment is crucial (Chiel & Beer, 1997). Assuming that all human-machine interactions are multisensory, our research objectives are (i) to better understand the role of a naturalistic multisensory environment on the orientation of attention during HCI, and (ii) to develop an instrument for real-world HCI tasks to measure the orientation of attention.

To accomplish these objectives, we conduct two laboratory studies examining the use of task-incongruent and task-congruent, but goal-unrelated, sensory perturbations to infer attentional orienting neurophysiological responses in naturalistic multisensory digital/physical environments using electroencephalography. To elucidate the role of naturalistic environments, multisensory integration, and attention during HCI tasks, we build on a conceptual framework that bridges multisensory integration mechanisms with attention (Talsma et al., 2010). This framework demonstrates how multisensory integration is sensitive to both bottom-up and top-down attentional forces. Consequently, we develop a multisensory perturbation technique to trigger a drift of the allocation of attention from the HCI task toward a distractor to measure the orientation of attention. Finally, utilizing multisensory microworlds, we iteratively increase naturalness to ensure that our measurement generalizes to quasi-naturalistic environments.

We empirically demonstrate that our approach is robust to complex multisensory environments while retaining sensitivity to task properties and top-down attention (i.e., attentional demand). We show that the naturalistic environment influences the neurophysiological response to perturbation while still being sensitive to top-down attentional mechanisms. Therefore, we present an approach that can be used to explore two significant IS and HCI research opportunities: (i) to understand better and measure attentional mechanisms in HCI, and (ii) to provide an approach that targets the impact of technology usage and design on relevant users' cognitive mechanisms (Dimoka et al., 2011; vom Brocke et al., 2020).

The structure of this paper is as follows. First, we present the conceptual background. Next, we provide an overview of the research, followed by a description of the procedures, methods, perturbation design, and results for Study 1 and Study 2. Finally, we discuss the implications of our findings.

4.2 Conceptual Background

4.2.1 Multisensory integration

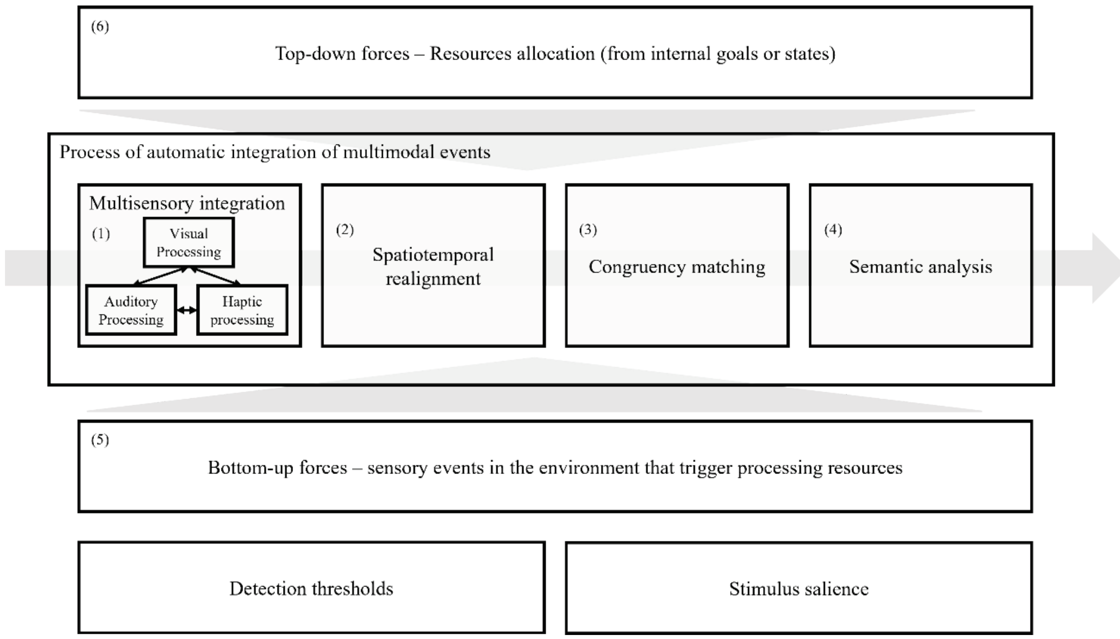
Humans possess multiple sensory systems that provide complementary impressions of the environment, which is essential for perceptual mechanisms, cognitive processing, and control of motor actions in an automated fashion (Meredith, 2002). Multisensory modalities such as olfactory, auditory, visual, or tactile systems support daily living perceptions of the environment. The brain has the ability to integrate complex and disparate information from different modalities into a unique and coherent perceptual experience of a multisensory event (Talsma et al., 2010).

Multisensory integration refers to “the set of processes by which information arriving from the individual sensory modalities (e.g., vision, audition, touch) interacts and influences processing in other sensory modalities, including how these sensory inputs are combined to yield a unified perceptual experience of multisensory events” (Talsma et al., 2010). While this process is automatic, there is evidence that it is intertwined with attention (Tang et al., 2016). For example, (Talsma & Woldorff, 2005) showed that expected multimodal stimulations of the senses increased early neurophysiological responses compared to unexpected stimulations.

Brain mechanisms integrate this information and even favor further downstream processing to positively bias or enhance the processing of multisensory events. Multiple sensory modalities increase the probability and speed of event detection events compared to single modalities (Diederich & Colonius, 2004; Gottfried & Dolan, 2003; Hagmann & Russo, 2016; Innes & Otto, 2019). It increases sensitivity and could improve perceptual abilities. Events stimulating multiple modalities have a higher chance of being prioritized for further processing, thus capturing human attentional resources and increasing detection time (Diederich & Colonius, 2004). This cognitive phenomenon is called the redundant signal effect.

Figure 29

Conceptual framework on the interplay between multisensory integration and attention



Note. Adapted from Talsma et al., 2010

The multisensory information processing depends on several essential conditions, as outlined in Figure 29. (1) A multisensory event is registered when its saliency is above the detection threshold in one modality (e.g., vision for a spatial task) (Talsma et al., 2010). If this condition is met, the brain will attempt to align the other modalities composing that event that are less dominant than the initial ones (e.g., auditory for the same spatial task). (2) Spatiotemporal realignment involves the alignment of sensory inputs across time and space (Talsma et al., 2010). Temporal and spatial alignment of sensory inputs is essential for the optimal integration of a multisensory event (Talsma et al., 2010). Aligned sensory inputs in different modalities have a higher chance of being prioritized for further processing, thus attracting human's attention (Driver, 1996; Van der Burg et al., 2008, 2009). When this condition is met, the perceived modalities are congruent and evoke the same event. In addition, (3) if the realignment and congruency between the different modalities match, the brain allocates more resources for further downstream processing (Talsma et al., 2010). Congruency is achieved when spatial characteristics and temporal features corresponding to multisensory events match. Examples of incongruent multisensory events are the McGurk illusion and the ventriloquist effects, in which lip movements do not match the auditory speech (Bertelson et al., 2000; McGurk & MacDonald, 1976). (4) Semantic analysis corresponds to high-level functions that can influence the integration of multisensory events, such as top-down processes (Talsma et al., 2010).

This process leads to the unity assumption (Welch, 1999), "which corresponds to the degree to which observers infer (consciously or not) that two sensory inputs originate from a single common cause". However, different modalities do not participate in the same degree of integration based on a bottom-up (stimulus-driven) process. The modality appropriateness hypothesis (Welch & Warren, 1980) posits that there are dominant sensory inputs depending on the characteristics of the task at hand. It means the task has a non-negligible influence on sensory inputs, integration, and processing.

In the field of HCI, multisensory integration plays a crucial role in the design of interactive experiences. In virtual reality environments, integrating multiple sensory inputs (e.g., auditory, visual, haptic) can lead to more engaging and immersive experiences for the user. Marucci et al. (2021) showed that during a driving task in virtual reality environments, visual-audio and visual-audio-haptic feedback enhanced the sense of presence compared to visual feedback alone. However, the feedback design does not need to be congruent with the task and can serve as an information vector. Cooper et al. (2018) tested the effects of "substitute multisensory cues" on performance and the sense of presence and immersion. They showed that multimodal feedback improved task performance and increased the perceived sense of presence compared to bimodal and unimodal cues. Nevertheless, poorly designed multisensory environments have drawbacks. Multisensory integration can lead to unattended effects. A mismatch of sensory signals providing feedback on the body orientation might cause cybersickness (Gallagher & Ferrè, 2018).

However, the characteristics of multisensory events are not the only factors influencing their integration. In addition, (6) top-down attentional forces influence the integration of multisensory inputs (Talsma et al., 2008; Talsma et al., 2010). Stimuli attract attention not only because of their inherent salience but also because of their relevance to current goals. Top-down attention, also referred to as endogenous attention, is the “processing resources allocated according to internal goals or states of the observer” (Talsma et al., 2010). In other words, top-down attentional forces refer to the attention that the individuals consciously or unconsciously direct toward the stimuli. It includes the volitional and executive control of attentional allocation, but also the context, learning, expectation, and historical aspects of the brain (e.g., selection history) during a task that influences perceptual processes (Gaspelin & Luck, 2018).

Top-down attention modulates multisensory integration and is already underway in the brain before a bottom-up signal arrives. Endogenous attention can be voluntarily allocated to a stimulus, a sensory modality, or a specific region of space in order to achieve task goals. Therefore, this mechanism is crucial for allocating attentional resources toward relevant stimuli and inhibiting irrelevant stimuli, modalities, and spatial locations (Tang et al., 2016). Evidence showed that directing attention to an attended sensory cue enhances sensory responses (Choi et al., 2018; Hillyard et al., 1973). Furthermore, attention can be oriented to a specific sensory modality based on internal goals (Tang et al., 2016). In this case, there is evidence for the presence of sensory gating mechanisms of the unattended modality (Talsma et al., 2007) and increased behavioral and neurophysiological responses to the intentionally attended modality (Talsma & Woldorff, 2005) (see Tang et al. (2016) for a comprehensive review on the interaction of endogenous and exogenous attention with multisensory integration).

To conclude, attention-grabbing events (bottom-up forces) and the executive function of attention (top-down forces) influence attentional mechanisms in the environment. This indicates that the relationship between attention and multisensory integration is situated at the intersection of the user, the task, and the environment (Macaluso et al., 2016).

4.2.2 Attentional Orienting and Perturbation

When faced with a multitude of stimuli, it is often impossible to process all the information simultaneously. Fortunately, the orientation of attention prioritizes processing certain stimuli or stimulus aspects at the cost of dealing less efficiently with others. Attentional orienting is a reflex that enables humans to immediately integrate cues in their environment under the form of multisensory stimulations (Sokolov, 1963). Attentional Orienting corresponds to "the process responsible for moving focus of attention from one location, feature or object, to another" (Talsma et al., 2010, p. 401). It corresponds to the automatic and covert task-directed mechanism involved in selecting relevant and inhibiting irrelevant sensory modalities, events, and task-related cues for further processing (Talsma et al., 2010). Attentional orienting is not necessarily followed by the detection, the conscious awareness, of the perturbation (Mulckhuyse & Theeuwes, 2010).

Therefore, attention can also be involuntarily captured by exogenous sensory stimulations, even if the event is unrelated to the current task (Öhman et al., 2001; Zhang et al., 2012). The allocation of resources to this shift of attention toward a non-goal-related perturbation can serve as a proxy measurement of the attentional orienting toward the task and is sensible to the multisensorial environment (Marucci et al., 2021; Zink et al., 2016), task properties (Burns & Fairclough, 2015; Marucci et al., 2021) and top-down forces (Horváth et al., 2008).

A perturbation is an unexpected exogenous stimulation from the external world (Tang et al., 2016). It can take the form of an auditory (Burns & Fairclough, 2015; Conrad & Newman, 2021; Ladouce et al., 2019; Zink et al., 2016), visual (Barutchu & Spence, 2021), somatosensory (Forschack et al., 2017), or multimodal stimulation (Bolton, 2015; Varghese et al., 2017). In this case, a perturbation is non-goal related and is triggered repeatedly with random intervals to avoid habituation and cognitive preparation due to its unpredictability (Allison & Polich, 2008; Duncan et al., 2009; Polich, 1989). This paradigm assumes that before a distracting event occurs, resources are allocated to optimize performance in the current task and that they are assigned to the processing of goal-related and task-congruent (semantically matching) stimuli that are likely to appear.

This orientation of attention to task-congruent features can further increase sensory integration and the resources allocated for its processing (Choi et al., 2018) and might inhibit non-relevant events. Moreover, as multisensory events are more salient than unisensory ones, attentional resources might be captured by the multisensory environment and gated for unisensory stimulation.

In EEG research, exogenous attentional orienting toward a stimulus generates an event-related potential (ERP). This neurophysiological response to a stimulus can characterize cognitive processes. In the case of sensory integration, a multisensory perturbation often presents an occipital P1 (positive component) and a frontocentral N1 response (negative component) (Talsma & Woldorff, 2005). Attentional resources toward a multisensory perturbation could be reflected by an enhancement of N1, followed by a late processing negativity in the frontal-central area. Multisensory perturbation involving body movement, in addition to auditory and visual stimulation, shows late potentials after N1 such as P2 and N2 (Varghese et al., 2017). These late potentials are sensitive to the availability of attentional resources for an unattended perturbation (Talsma et al., 2007). In the case of multisensory perturbation, Quant et al. (2005) hypothesized that late components (P2, N2) could denote the cognitive processes linked with the task demand and the processing of the perturbation.

Applied to HCI environments, multisensory aspects of the environment could influence the resources allocated to the attentional orientation toward an exogenous perturbation. A multisensory HCI environment could be attention-grabbing, which takes the form of directed attention resources toward processing information and inhibiting unrelated events. As discussed previously, it can represent the cognitive mechanisms that could lead to an increased sense of presence or immersion for the user (Marucci et al., 2021). Thus, we posit that a multisensory HCI environment will capture attentional resources and be observable under inhibited late components in ERPs triggered by a perturbation.

H1: High multisensory HCI environment will reduce the resources allocated to processing unexpected perturbations compared to a low multisensory one.

However, not only the properties of the environment influence attentional mechanisms. Attentional orienting is sensitive to the bottom-up multisensory aspect of the perturbation, the environment, and top-down forces (Talsma et al., 2010). As ERPs, neurophysiological responses to this shift of attention are also sensitive to the user's internal state. The properties of the tasks could require an increase in attentional demand and thus force the user to voluntarily orient attentional resources toward the task or specific aspect of it. In this case, the user directs the top-down orientation of attention (motivation, concentration) when necessary and can be induced by the task demand. It can be applied when task-related feedback is necessary to solve the task.

The sensitivity of early components, particularly the N1 evoked potentials (the negative amplitude around 100 ms), has been linked to top-down attentional processes. Quant et al. (2004) demonstrated that the N1 amplitude elicited by multisensory perturbation is reduced during a cognitively demanding task. The reduced magnitude of N1 may suggest the influence of top-down allocation of attentional resources toward the task on the early cortical activity following a perturbation. The inhibition of N1 may be attributed to the attentional demand required to process sensory information about the unexpected event. Similarly, multisensory events consisting of temporally and spatially aligned visual and auditory sensory perturbations have been shown to be enhanced with voluntary attention but reduced for ignored perturbations (Hopfinger & West, 2006; Talsma & Woldorff, 2005). These results show that the N1 component is modulated by a gating mechanism in the early perceptual stage that can be controlled by top-down attention.

Applied to HCI, aspects of the computer task could influence the attentional resources required to achieve goals. Attentional demand requires the voluntary allocation of attentional resources to the task. For example, while investigating the cognitive mechanisms of immersion, Burns and Fairclough (2015) showed that ERPs components were influenced by a demand increase while playing a computer game using an irrelevant unisensory perturbation. Terkildsen and Makransky (2019) showed that early component N1 amplitude was significantly lower during a high-presence task than a low-presence task during a computer game. It is to be noted that all the evidence above uses irrelevant auditory stimulations, and none of them used environmentally congruent perturbations. However, and most importantly for this research, this evidence shows that early components of ERPs are sensitive to top-down attention from the user, and the measure is sensible even in complex environments and tasks such as HCI. Thus, we posit as a second hypothesis:

H2: High attentional demand periods (top-down attentional orienting toward the multisensory HCI environment) will reduce the resources allocated to processing unexpected perturbations compared to low attentional demand.

4.3 Research Overview

To test the hypotheses, we designed two experiments aimed at developing perturbations to measure the allocation of resources to a shift of attention to a non-goal-related event. In the first experiment, we manipulated the sensory modalities of the environment to assess the effect of environmental naturalness on attentional orientation, thereby testing H1.

In the second experiment, we similarly manipulated the sensory modalities of the environment. However, we also adjust the attentional demand of the task to measure the sensitivity of the neurophysiological response, even in naturalistic environments and tasks, to top-down mechanisms. In doing so, we test both H1 and H2.

Table 30

Overview of the Experiments

Experiment	Methodology	Hypothesis	Perturbation	
1	Experiment manipulating multisensory HCI environment	H1	Sensory Modality	Auditory
			Task/Environment Relevance	Task-irrelevant (exogenous distractor):
			Task/Environment Congruence	Incongruent (unimodal synthetic tone)
2	Experiment manipulating multisensory HCI environment and attention demand	H1 and H2	Sensory Modality:	Multimodal
			Task/Environment Relevance	Task-irrelevant (exogenous distractor)
			Task/Environment Congruence	Congruent (multisensory perturbation)

4.4 Study 1 - Multisensory HCI Environment and Attentional Orienting

In study #1, we test H1: a high multisensory HCI environment will show a reduction of the resources allocated to processing unexpected perturbations compared to a low multisensory one. We manipulated the naturalness of the environment with two levels, a high multisensory environment corresponding to the spatiotemporal alignment of three modalities (motion, visual, auditory) against a low multisensory environment (visual, auditory). All feedback modalities were temporally aligned with the task; here, a driving game.

4.4.1 Procedure and manipulation

D-BOX Technologies Inc. (Longueuil, Canada) designed and programmed the multisensory HCI environment stimuli, which were delivered via a D-BOX VK seat. The motion modality was embedded in a proprietary motion code file that was designed to synchronize with the computer videogame. The videogame was a driving game named “Dirt: Showdown” produced by Codemaster (Warwickshire, UK).

We used a between-group design that manipulated the HCI environment's multisensory aspect. At their arrival, the participant was divided into two groups, either in the high multisensory (HMS) condition (i.e., auditory, visual, motion) or low multisensory (LMS) (i.e., auditory, visual) condition.

The task was composed of a first race, so the participants familiarized themselves with the controls. This first training race was the same for every participant with respect to their condition. Then, four races were randomly performed, and all participants played the same four races; only the order differed. To avoid disturbing the participant during the task and maintain immersion within the task some validity concerning the, the race order was given at the start. Then, the participant was instructed to autonomously do the race following a printed order provided to the participant.

As a perceived measures and manipulation check, the participant filled out the Immersive Experience Questionnaire (IEQ), a subjective gaming experience questionnaire composed of 31 items (Jennett et al., 2008). The questionnaire was administered via an iPad at the end of each race.

4.4.2 Perturbation

To elicit attentional orientating toward a distractor, an auditory perturbation was designed. The perturbation duration was 1000 Hz sinewave tone of 100 milliseconds, with a sample rate of 44 100-kHz, and a 32-bit depth. The perturbation was presented just behind the participants at head level. The auditory perturbation was played with a second Mission 761 speaker placed 30 cm behind the participants at head level. The auditory perturbation was parametrized to be, on average, 10 dB higher than the game in order to be heard. We measured an average loudness of 84.2 dB for the tone (forward measurement = 84.2 dB, backward measurement = 85.5 dB). Loudness was measured using a REED R8080 sonometer setup at head level. This was particularly important to ensure the salience of the perturbation. The auditory stimuli were triggered randomly during the task using an interstimulus interval (ISI) following a gaussian distribution with an average of 12 seconds and a standard deviation of 3 seconds to avoid habituation and predictability.

To measure the onset of the auditory perturbation in the EEG data, we designed the auditory sound to be 2.0 stereo audio file with the right channel being the perturbation (100 ms, 1000 Hz sinewave tone), the left one was a square wave pulse (5 ms, 1000 Hz square pulse) that was rerouted directly from the receiver to the trigger box it was directly transmitted and marked as an event in the EEG recording.

The audio channels of the game were routed via HDMI through a Pioneer VSX-324 AV receiver, with the front left and front right channels played on Mission 761 speakers (Mission, Huntingdon Cambridgeshire, UK) at a mean loudness of approximately 75.4 dB (forward measurement = 75.4 dB, backward measurement = 73.9).

4.4.3 Statistics

Sensor-level EEG statistical analysis is performed with cluster-based permutation tests following (Maris & Oostenveld, 2007) in Python (3.8.11) and MNE (1.3) (Gramfort et al., 2013). In study #1, we developed a between-subject one-way design and performed a spatiotemporal cluster-based permutation F-test on all the sensor data. In study #2, we developed a within-subject two-factor design, where we applied a spatiotemporal cluster-based permutation repeated measures ANOVA. All statistics are applied at the 2nd level, the averaged epochs of the evoked EEG responses. The FieldTrip neighbor templates are used to compute the adjacency between sensors. We applied the tests with 1000 permutations on all sensor (32 channels) data in an interval between -100 ms and 800 ms relative to stimuli onset. Critical F-value thresholds for forming a cluster were computed a priori, based on the statistical test employed, the number of participants, directionality, and the design, and for an α value of 0.05.

The remaining statistical analyses, behavioral and subjective, are performed in R (4.1.0). Analysis of variance (ANOVA) is used in both study #1 (between-group) and #2 (within-subject). To assess the internal constancy of the psychometric instruments, we computed Cronbach's α with a bootstrap confidence interval (CI) based on 1000 samples at 95%. In addition, heteroscedasticity (non-constant error variance) and the normal distribution of the residuals are checked. When applicable, we performed a secondary analysis where statistical tests were non-significant. We calculate Base Factors (BF) with default priors to evaluate the absence of effect.

4.4.4 Participants

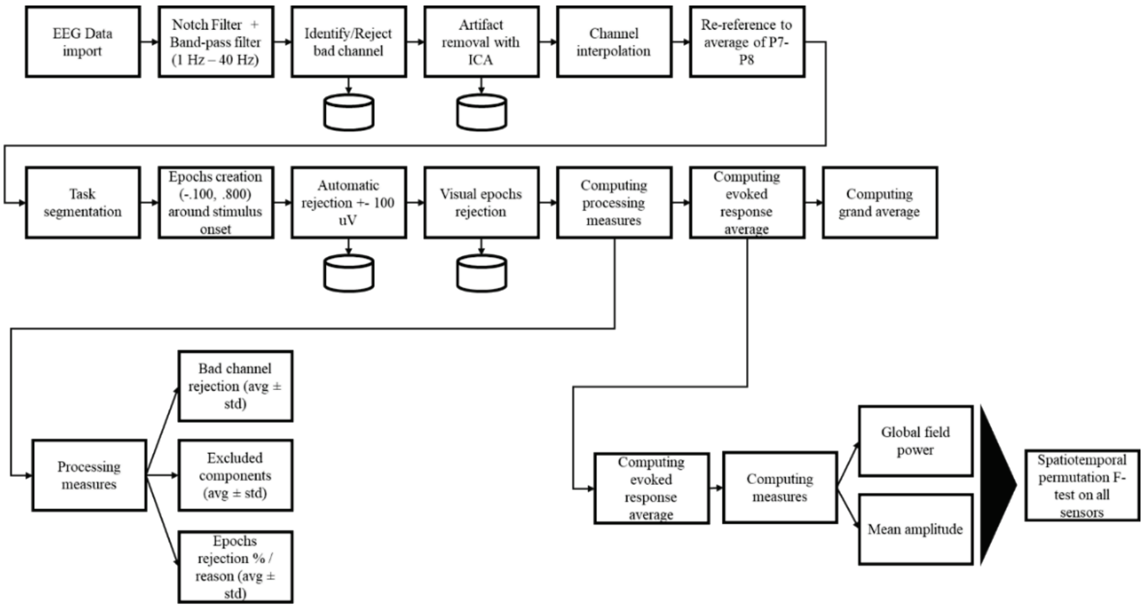
Sixteen healthy volunteers (mean = 26.65 ± SD 6.44 years old, 5 females) took part in the study. Participants were screened based on good health and normal to corrected vision. The experimental certificate was approved by the ethics committee of our institution (2022-4686). All participants gave their written and signed consent before the experiment. Participants were compensated 50 CAD after their participation in the study.

4.4.5 EEG processing

EEG signals were recorded raw at a 1000 Hz sampling frequency using BrainVision Recorder (Brain Products GmbH). EEG data processing was conducted with MNE (Gramfort et al., 2013) running on Python 3.7 (Figure 30). A band-pass filter (1 Hz – 40 Hz) was applied with a notch filter at 60 Hz. Filter parameters were the following: Butterworth 2nd order IIR filter, hamming window with 0.0194 passband ripple and 53 dB stopband attenuation, lower transition bandwidth: 1.00 Hz (-6 dB cutoff frequency: 0.50 Hz), upper transition bandwidth: 10.00 Hz (-6 dB cutoff frequency: 45.00 Hz). Then, we conducted artifact removal using Independent Analysis (ICA) with Infomax method. Bad channels were removed before ICA and interpolated with a spherical spline interpolation after. Bad epochs were first automatically removed with a threshold of +-100 uV on the signal amplitude. Visual scanning for "sanity check" and rejection was then performed.

Figure 30

EEG processing pipeline



The processing of the EEG showed no wide difference between our conditions in the quality of the data (Table 31). The condition without motion shows relatively similar compared to the condition with motion in terms of bad channel removal (Low = 1.13 ± 1.13; High = 1.50 ± 1.20) or excluded components due to artifacts (Low = 3.00 ± 0.76; High = 3.25 ± 0.89). We observed a slight difference in the number of epochs automatically rejected with a threshold of +- 100 uV (Low = 6.50 ± 5.73; High = 8.75 ± 7.40), but that was expected has the activation of the chair-generated noise due to the motion.

Before computing the grand average and the statistical analysis number of trials and the corresponding epochs were equalized. The approach tries to select for analysis an equal number of cleaned epochs occurring as close as possible in time between the two conditions. The argument for applying this technique is that we can assume a number of time-varying confounds in the epochs resulting from repeated auditory stimulation (e.g., noise, habituation). Thus, this method reduces this effect by minimizing the differences in the times of the events between participants.

Table 31

Descriptive statistics of the EEG processing pipeline

	Low multisensory HCI environment	High multisensory HCI environment
Task average duration	853.14 ± 22.97	869.06 ± 16.62
Bad channel	1.13 ± 1.13	1.50 ± 1.20
ICA – Excluded Components	3.00 ± 0.76	3.25 ± 0.89
# of epochs	61.63 ± 2.07	59.38 ± 2.83
# of automatically rejected epochs (+- 100 uV)	6.50 ± 5.73	8.75 ± 7.40
# of epochs after auto rejection	55.25 ± 7.55 (15.81 % ± 16.33 %)	51.25 ± 6.78 (14.43 % ± 11.69 %)
# of manually rejected epochs	6.50 ± 2.45	6.38 ± 3.11
Final # of epochs	48.75 ± 7.57 (21.25 % ± 10.61 %)	44.50 ± 6.23 (25.29 % ± 10.46 %)

Note. Mean ± std (% of total # of epochs)

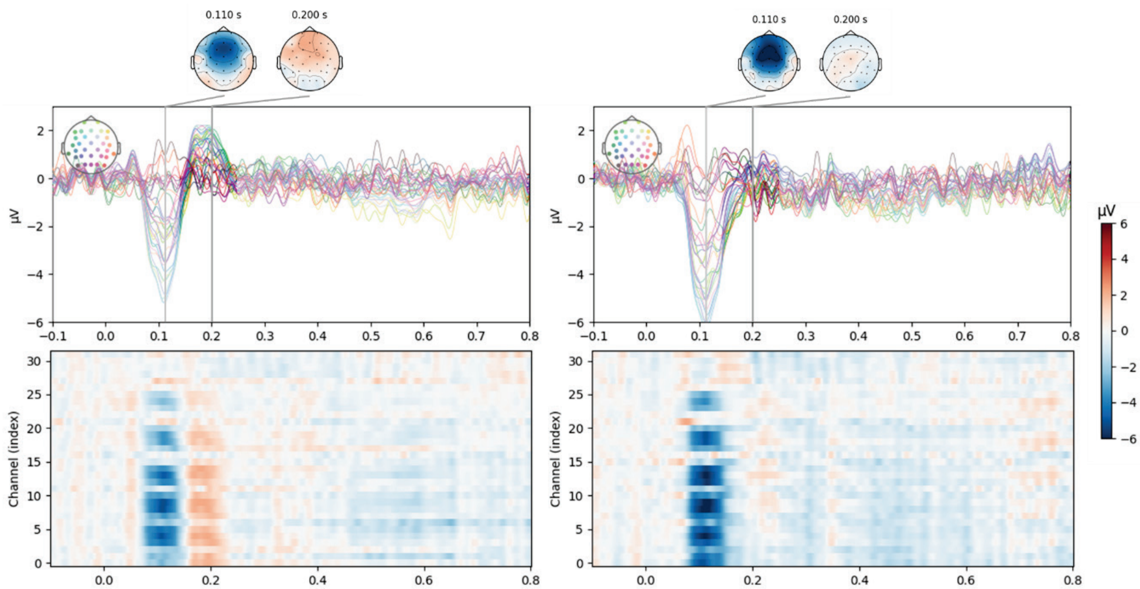
4.4.6 Results

4.4.6.1 Electroencephalography

Grand Average ERPs are plotted in Figure 31. A visual analysis of the ERPs shows suppression of the positive component (P200) around 200 ms just after the first exogenous components (N100).

Figure 31

Grand average of the auditory ERP

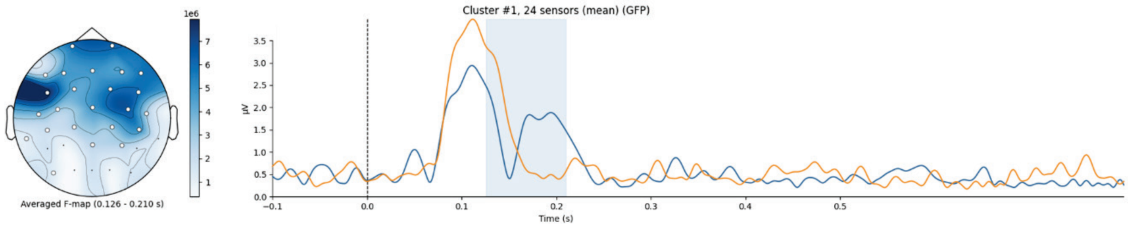


Note. Left = low multisensory HCI environment, right = high multisensory HCI environment

To test the difference between our two conditions (high and low multisensory HCI environment), we ran a between-condition spatiotemporal permutation one-tailed F-test on all sensors following (Maris & Oostenveld, 2007). The F threshold was selected a priori for a p-value of 0.05 for the given number of observations (Threshold = 4.600110, alpha level = 0.05).

Figure 32

Statistical topography of the mean F-values



Note. The color tone indicates the effect size. Blue = low multisensory HCI environment, Orange = high multisensory HCI environment

The statistical test result shows a significant difference (p-value = 0.031) in the effect of the two conditions on ERP in frontal and central areas between 126 and 210 milliseconds after initiating the sound stimulus (Figure 32), supporting H1.

We extracted the average amplitude at P2 for the sensor Fz as the midline response of ERPs was shown to be the strongest in our data, and it is in line with the relevant literature. LMS condition presents an average amplitude of P2 at Fz of -0.02 microvolt (SD = 2.09, range: [-1.88, 4.56]), while HMS condition shows an average amplitude at the same location 1.99 mv (SD = 1.20, range: [-0.25, 3.51]). The ANOVA suggests that the main effect of the multisensory HCI environment is statistically significant: F(1, 13) = 5.03, p = 0.043; ηp2 = 0.28, 95% CI [0.06, 1.00]).

4.4.6.2 Self-perceived measure

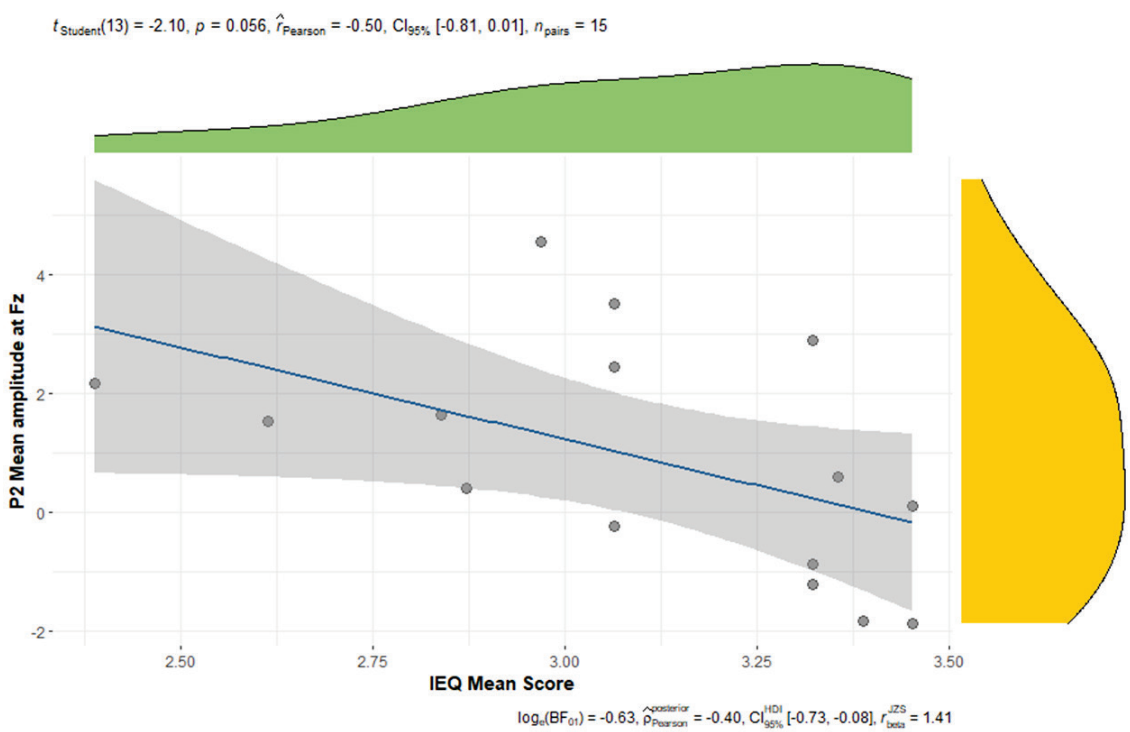
We computed Cronbach's α for internal constancy of the psychometric instrument IEQ (Items = 31) with a bootstrap at 95% confidence interval (CI) based on 1000 samples. The result shows Cronbach's α = 0.766, CI = [0.416, 0.871]. The ANOVA suggests that the main effect of MS HCI is statistically significant and large (F(1, 13) = 6.55, p = 0.024; η_p^2 = 0.33, 95% CI [0.03, 1.00]).

4.4.6.3 Post-hoc analysis

To look at the relationship between P2 amplitudes and self-perceived immersion, we correlated P2 Fz amplitude with the IEQ aggregated score. The Pearson's correlation test revealed that P2 amplitudes at Fz were negatively correlated with IEQ Mean Score across subjects, but this effect was not statistically significant (p=0.056). However, the effect size (r=-0.50) is considered large per Cohen's (1988) conventions (Figure 5.). The Bayes Factor for the same analysis revealed that the data were 1.881 times more probable under the alternative hypothesis as compared to the null hypothesis. The BF can be considered anecdotal evidence (Jeffreys, 1961) in favor of the alternative hypothesis (the existence of a correlation between P2 amplitude and subjective immersion).

Figure 33

Relationship between P2 mean amplitude and self-perceived immersion



4.5 Study 2 - Multisensory HCI Environment, Attentional Demand, and Attentional Orienting

In study #2, we test H1 (a high multisensory HCI environment will show a reduction of the resources allocated to processing unexpected perturbations compared to a low multisensory one). Furthermore, we test H2 (high attentional demand periods (top-down attentional orienting toward the multisensory HCI environment), which states that we should observe a reduction of the resources allocated to processing unexpected perturbations compared to low attentional demand ones.). We manipulated the naturalness of the environment in the same fashion as study one by adding a modality, motion. It further increases the simulated environment's ecological validity by offering the sensory feedback of the body in space in the simulation, an essential sensory modality in driving. All feedback modalities were also temporally aligned with the task; here, a realistic driving simulation. Attentional demand was manipulated based on the lap sectors (curves were identified as high-demand sectors, straight as low-demand sectors).

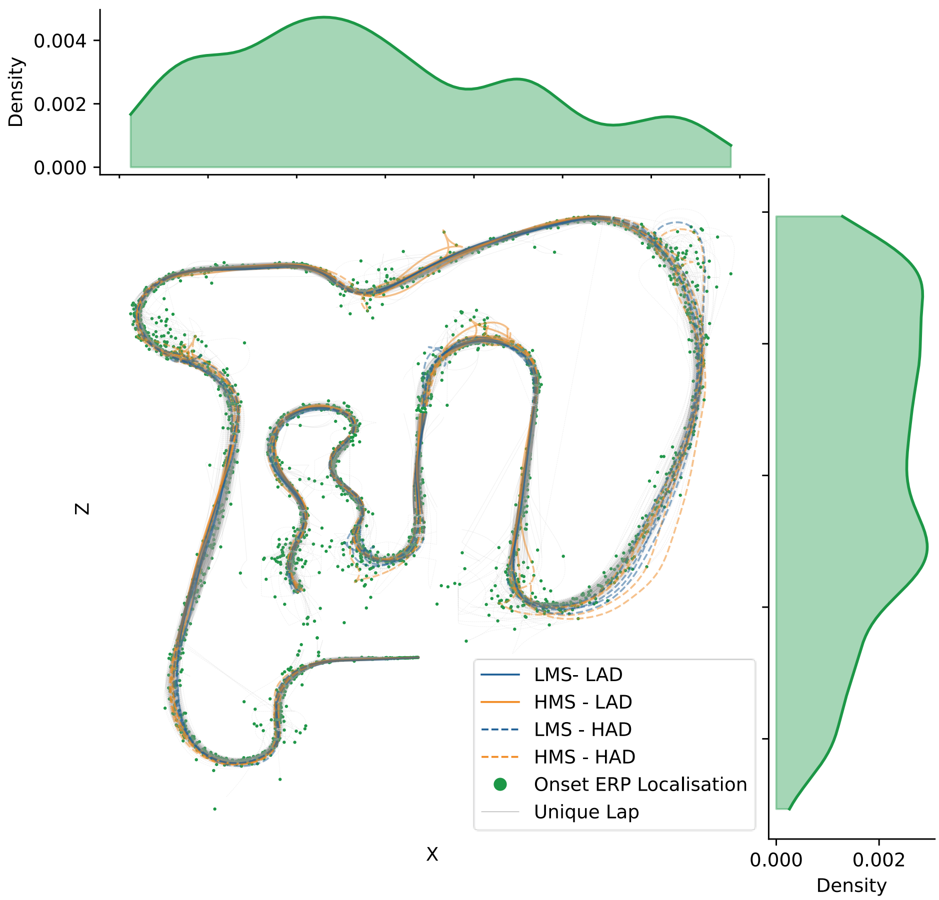
4.5.1 Procedure and manipulation

This study uses a within-subject design with two factors. The first factor corresponds to the level of the multisensory environment (motion, visual, auditory). The second factor is the attentional demand of the task and is derived from the telemetry data corresponding to the circuit sections, straights, and curves sections (Figure 34), as low attentional demand (LAD) and high attentional demand (HAD), respectively.

Participants were instructed to go the faster they could. They had two practice runs to discover the circuit and the controls. Then, they ran for ten laps. The simulator automatically stopped at the end of each lap, and the lap time was given. The participants were then instructed to complete a short questionnaire before starting the next run. Conditions were randomized between laps but balanced (50%).

Figure 34

Racing trajectories and localization of perturbations



Note. Racing Circuit trajectory corresponding to multiple runs from a random participant (blue and orange line). Green points represent the location of onset perturbation triggered during the laps. On top and left sides, the probability density function of the localization of the onset perturbation on the X and the Z axis.

The perturbation was designed to fit the multisensory environment (visual, sound, movement). It took the form of a short vertical perturbation of 200 ms with an amplitude of a maximum of 1 cm of the whole chair. The vertical perturbation was relative to its position when triggered to ensure saliency. Perturbations were irregularly triggered to avoid habituation and prediction from the participants. The interstimulus interval was randomly selected following a gaussian curve with an average of 7 seconds and a standard deviation of 2 seconds.

To measure the perturbation onset, we used the 3D Acceleration Sensor from Brain Products (Germany) set at a sampling frequency of 1000 Hz (dynamic range of ± 2 g, sensitivity of 1450 mV/g ± 10%) connected to BrainAmp ExG AUX Box. The orientation of the acceleration was recorded on three orthogonal axes (x, y, z): lateral, distal, and ventral. The hardware was attached to the back of the chair, with the y position measuring the vertical acceleration.

Data collection and signal synchronization were implemented using LabStreamingLayer protocol with LabRecorder as central recording software. EEG data were redirected from the amplifier to BrainAmp Series LSL connector. The LSL connector was used to stream the EEG stream, which was composed of 32 EEG channels and three accelerometer channels sampled at 1000Hz. Following advice from BrainVision, impedances, signal quality, and troubleshooting were conducted with BrainVision Recorder. BrainVision LSL Viewer was used to monitor the signal quality during the experiment.

Figure 35

Experimental setup and simulator environment

The driving simulator was eXpanSIM developed by Raving Bots (Wrocław, Poland). All telemetry data from the simulator is accessible through their C++ SDK. We integrated a data stream using LSL protocol as a plugin for the simulator using the SDK for this data collection; the plugin streams all the variables at a 50 Hz sampling rate in real-time. The data was recorded and synchronized using the same recording software as the EEG, accelerometer, and LabRecorder. It enables the telemetry data to synchronize with the neurophysiological data with a precision of milliseconds.

As a perceived measures and manipulation check, the participant filled out a questionnaire administered via an iPad at the end of each race. The perceived presence was composed of 3 items (Jennett et al., 2008; Marucci et al., 2021). The reduced number of items was selected to fasten data collection between each lap.

4.5.2 Participants

Twenty-two healthy male expert driver volunteers (mean = 36.28 ± SD = 10.74 years old) participated in the study. Participants were screened on the basis of good health and normal to corrected vision. All participants were selected with experience with race simulators and real-world sports driving. The experimental certificate was approved by the ethics committee of our institution (2022-4686). All participants gave their written and signed consent before the experiment.

4.5.3 EEG processing

The EEG signal recording and processing pipeline were precisely the same as in study #1. The raw signal was recorded at a 1000 Hz sampling frequency using LabRecorder , the central recording software based on LabStreamingLayer . A 32 electrodes headset was used following the international 10–20 system. EEG data processing was conducted with MNE (Gramfort et al., 2013) running on Python 3.7. A band-pass filter (1 Hz – 40 Hz) was applied with a notch filter at 60 Hz. Filter parameters were the following: Butterworth 2nd order IIR filter, hamming window with 0.0194 passband ripple and 53 dB stopband attenuation, lower transition bandwidth: 1.00 Hz (-6 dB cutoff frequency: 0.50 Hz), Upper transition bandwidth: 10.00 Hz (-6 dB cutoff frequency: 45.00 Hz). Then, we conducted artifact removal using Independent Analysis (ICA) with Infomax method. Bad channels were removed before ICA and interpolated with a spherical spline interpolation after. Bad epochs were first automatically removed with a threshold of +-100 uV on the signal amplitude. Visual scanning for "sanity checking" and rejection was then performed.

ERPs onset of the perturbation was computed via the accelerometer due to latency between the order sent to the chair and the actual physical movement of the car (~ 100 ms). Thus, the upward trajectory in the vertical axis y was used. To test our hypothesis on the effect of a multisensory HCI environment (i.e., audio-visual-motion / audio-visual) and the effect of attentional demands of the section, we used the synchronized telemetry data to map the exact location of the vehicle and section during which a perturbation has been generated.

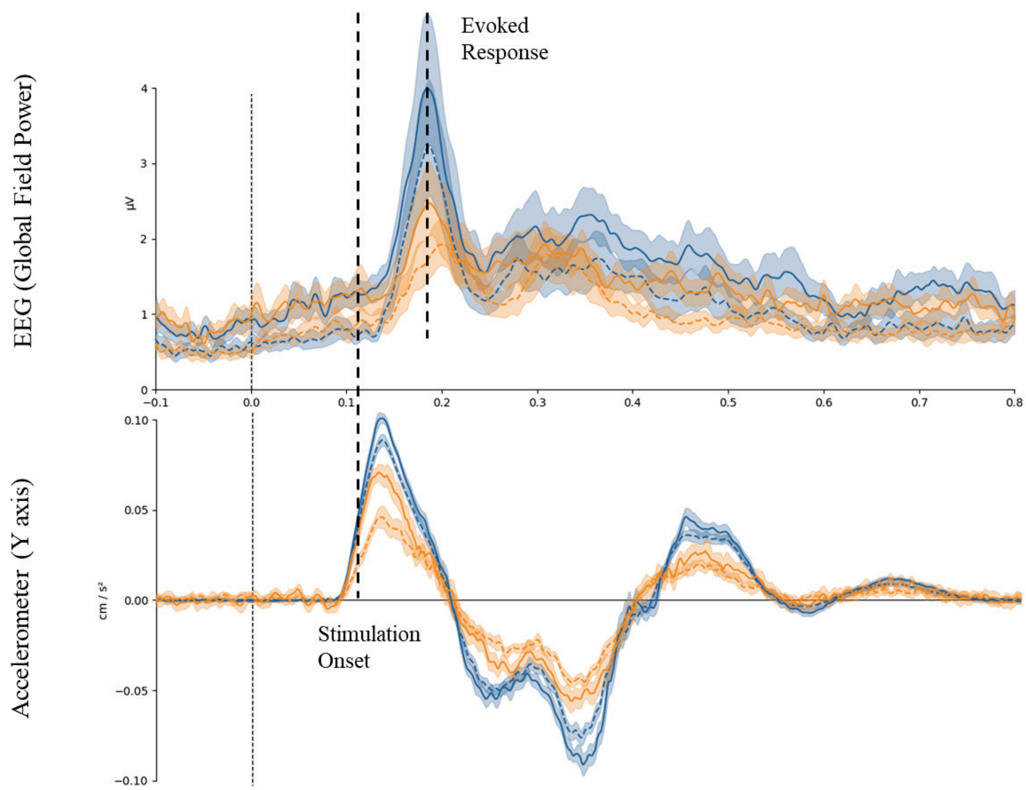
4.5.4 Results

4.5.4.1 Electroencephalography

Firstly, we computed the average perturbation measured via the accelerometer on the Y-axis to ensure that the multisensory events triggered a neurophysiological response. Next, we computed the averaged Global Field Power (GFP) for each condition to observe this response better and aligned the data with the accelerometer. The data shows that the perturbation precedes the neurophysiological response, indicating that the vertical stimulation provoked increased electrical activity measured at the scalp's surface (Figure 36).

Figure 36

ERP to the multisensorial perturbation

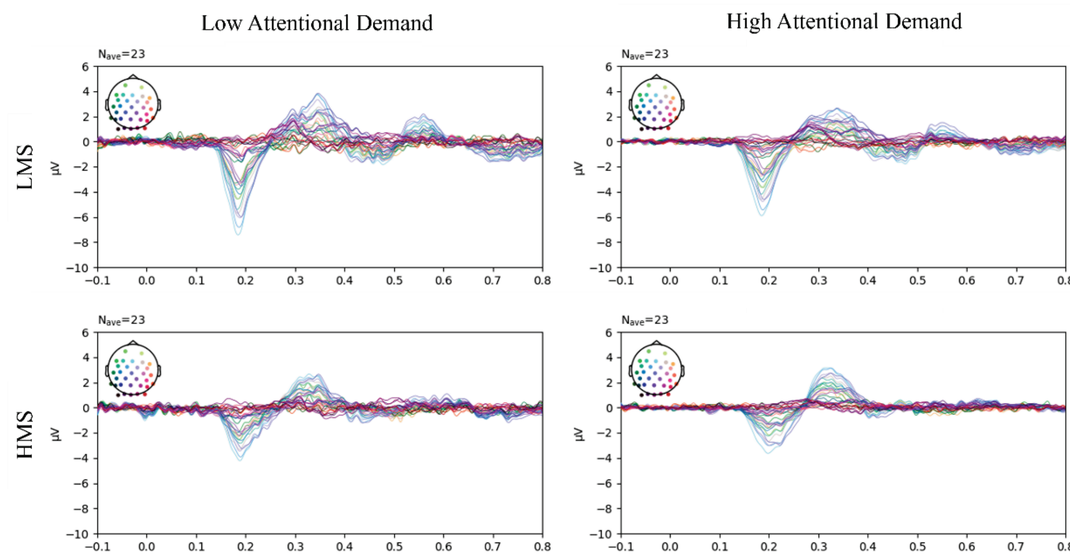


Note. On top, global field power represents all the electrodes' spatial standard deviation over time. Bottom: Accelerometer data in the Y-axis showing the perturbation. Blue = low multisensory HCI environment, Orange = high multisensory HCI environment. Full Line = Low demand, Dash line = high demand

Grand average ERPs obtained during the study are presented in Figure 37 to offer a visual indication of the neurophysiological response. While the same perturbation was triggered between conditions, substantial differences between the responses can be observed.

Figure 37

Grand average ERPs for each channel per conditions



Note. High multisensory (HMS) and Low multisensory (LMS) conditions

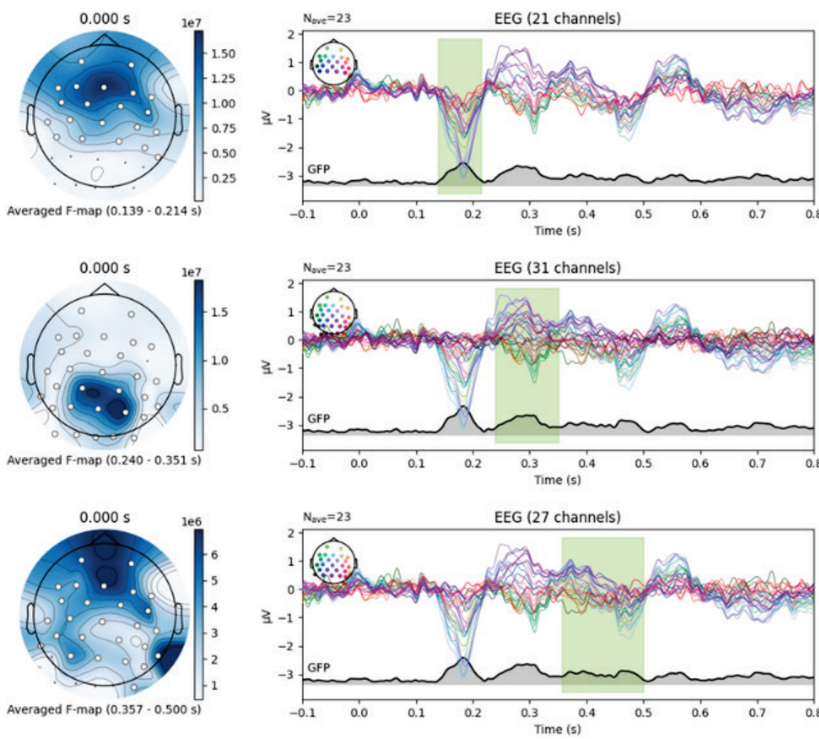
To test our hypotheses, we performed a two-way repeated measures ANOVA on sensor data with spatiotemporal clustering (Figure 38, 39). Critical F-value threshold for a two-way ANOVA (repeated measures, within factors, one-tailed) was computed for 23 participants, 2x2 conditions, and a p-value of 0.05. The a priori analysis gives us a F threshold = 4.300949.

The interaction effect was not significant, all clusters p-value were > 0.05. However, the cluster-based permutation test indicated a significant main effect between conditions low and high of the multisensory HCI Environment (Figure 38). Three clusters can be observed in the data between approximately 139 to 214 ms in the frontal area ($p = 0.005$), 240 to 351 ms parietal area ($p = 0.005$), and 357 to 500 ms in the whole scalp ($p = 0.011$).

The cluster-based permutation test indicated a significant main effect between conditions low and high attentional demands (Figure 39). One cluster can be observed between approximately 160 to 200 ms in the data (p = 0.0285), covering the frontal, central and parietal areas.

Figure 38

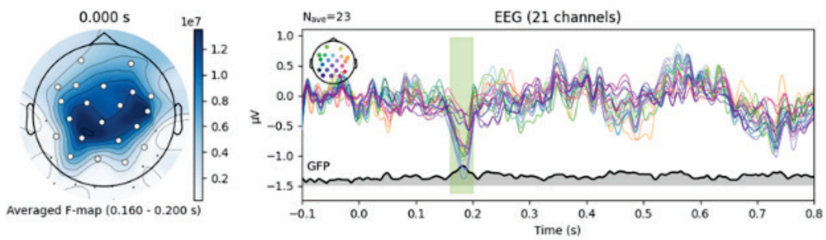
Main effect of Multisensory HCI Environment



Notes. The grand average ERP represents the subtractions of LMS by HMS.

Figure 39

Main effect of Attentional Demand (Low – High)



Notes. The grand average ERP represents the subtractions of LAD by HAD.

4.5.4.2 Self-Perceived Measure

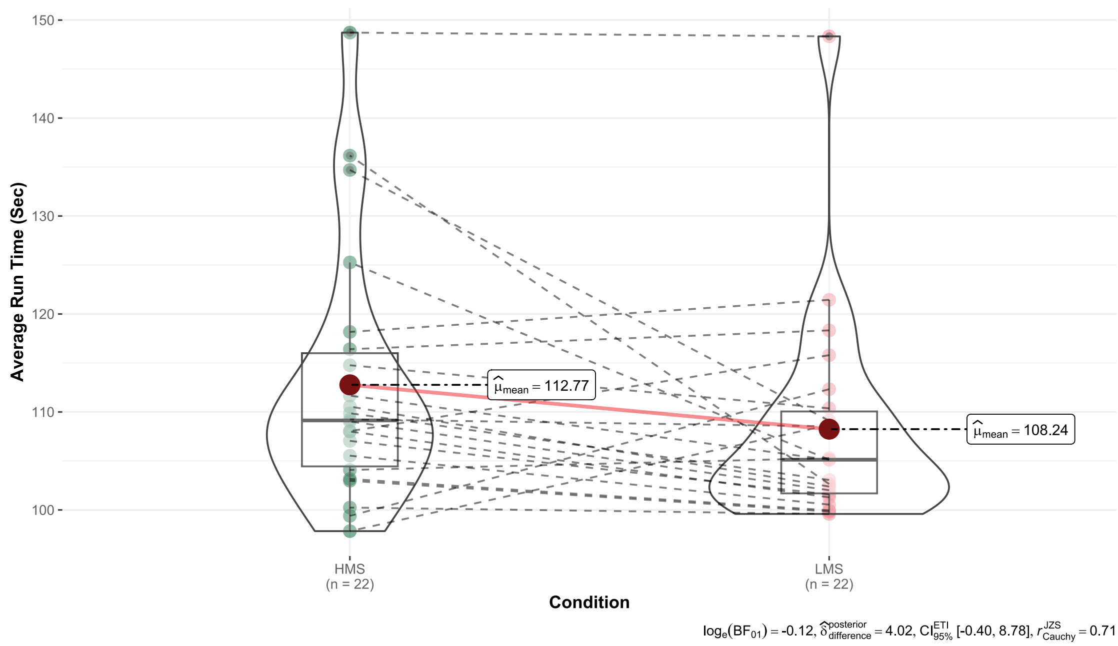
We computed the Cronbach's α for internal constancy of the psychometric instrument of Presence (Items = 3) with a bootstrap at 95% confidence interval (CI) based on 1000 samples. The result shows Cronbach's $\alpha = 0.806$, CI = [0.752, 0.850]. The average mean of perceived presence was 3.04 (SD = 2.02) for HMS and 2.77 (SD = 2.51) for LMS. An ANOVA suggests that the main effect of MS HCI environment is statistically insignificant $F(1, 31) = 1.27$, $p = 0.268$; $\eta_p^2 = 0.04$, 95% CI [0.00, 1.00]. The Bayes Factor for the same analysis shows a $BF_{10} = 0.338$, representing anecdotal evidence for the null hypothesis compared to the alternative hypothesis.

4.5.4.3 Behavioral Measure

We computed the performance as the average lap time between conditions, see Figure 40. The average run time is superior for HMS HCI condition with 112.77 seconds (SD = 12.94, range: [97.84, 148.72]) than the LMS condition with 108.24 seconds (SD = 10.89, range: [99.59, 148.34]). However, the difference is not statistically different. The repeated-measures ANOVA suggests that the main effect of MS HCI environment is statistically not significant ($F(1, 21) = 3.87$, $p = 0.062$; $\eta_p^2 = 0.16$, 95% CI [0.00, 1.00]). The Bayes Factor for the same analysis shows a $BF_{10} = 1.310$, representing anecdotal evidence for the alternative hypothesis compared to the null hypothesis.

Figure 40

Performance per conditions



Note. Performance measured as Average Run Time in seconds between low multisensory HCI environment (LMS) and high multisensory HCI environment (HMS).

4.6 Discussion

The result of this study offers several significant contributions to research in HCI on the role of multisensory integration and attentional orienting and its measurement. This research's general goal was to better understand the orientation of attentional mechanisms and their relationship with the multisensory aspect of real-world HCI tasks. Building on a conceptual framework bridging attention and multisensory integration (Talsma et al., 2010), we provide a methodological approach and evidence demonstrating the possibility of exploring those covert mechanisms even during complex and dynamic tasks in quasi-naturalistic environments using a perturbation technique.

We show the impact of naturalistic HCI environments on attentional orienting. Our results show that ERPs triggered by a shift of attention toward an auditory perturbation are influenced by multisensory environments (number of sensory modalities) of the task and environment. Supporting H1, we showed a reduction in attentional resources allocated to external auditory distractors in the high multisensory HCI condition compared to conditions with visual and audio. This observation reflects a presumed increased attentional capture of multimodal environment and gating of unrelated distractions (Talsma et al., 2010).

Study one showed that the early processing of stimuli is sensitive to attentional resources available. These results show a potential reduction of allocated attentional resources to the auditory distractors observable around 200 ms with a positive early component in the control condition, which is suppressed in the high multisensory HCI environment. The increased orientation and the reduced resources allocated to a shift of attention toward the auditory stimulation can indicate an increase in the orientation of attention toward the task imposed by the multisensory environment when a new sensory modality is added. In line with (Burns & Fairclough, 2015), this observation of the neurophysiological response also reflects a putative increase in attentional capture.

Using a non-task-relevant auditory stimulation, we measured the orientation of attention by the resources allocated to a perturbation. However, to reduce some limitations of study #1, in study #2, we designed the perturbation to be task-relevant and fit the multisensorial environment using a vertical upward perturbation of the chair. Auditory stimulation might not respect the modality appropriateness hypothesis (Talsma et al., 2010), thus inducing a more significant reduction in the shift of attention toward the distractor due to the saliency of the environment.

Table 32

Study, perturbation, and hypothesis summary

	Study 1	Study 2
Perturbation parameters	Modality: Auditory Task Relevance: Task-irrelevant (exogenous distractor) Task/Environment Congruence: Incongruent (synthetic tone)	Modality: Multimodal Task relevant: Task-irrelevant (exogenous distractor) Task/Environment Congruence: Congruent (multisensory upward perturbation)
H1 (High-Low Multisensory Environment)	Supported	Supported
H2 (High-Load Attentional Demand)		Supported

In study two, we manipulated the naturalness of the multisensory environment to ensure the sensibility of the neurophysiological response of our novel perturbation design (H1). In the same manner, the objective of this study is to observe and measure the orientation of attention by the resources allocated to a perturbation. However, this distractor is a multisensory stimulation that fits the context of driving in a multimodal simulator. Similar to study one, we hypothesized that a high multisensory HCI environment would reduce the resources allocated to processing unexpected perturbations compared to a low multisensory one. We also manipulated the attentional demand based on the properties of specific task sections (H2). Some sections required additional top-down attention orientation, impacting unexpected distractor processing.

The results show the presence of a gating mechanism applied to the perturbation within a high multisensory HCI environment. In line with the orientation of attention, the brain favors useful sensory feedback necessary for the task and inhibits irrelevant distraction, even in the same sensory modalities. This effect was observed in the neurophysiological response in the ERPs with an inhibition of the late components after N1 (Varghese et al., 2017). This explanation is in line with Quant et al. (2005) hypothesis that late components of multisensory perturbation might be related to task demands or cognitive processing. Thus, the multisensory environment may indeed increase the orientation of attention toward the task while reducing the resources allocated to processing the perturbation.

Subjects allocated fewer resources to the distractor during the conditions in high multisensory environments and the demanding sections of the task. N100 appears to influence the subject's top-down attention, showing the effect of task demand on top of the multisensory environment. This observation aligns with current research showing that early ERP responses to multisensory perturbation (i.e., audio, visual, and body motion) are endogenous components and sensitive to the internal state of the subject (Quant et al., 2004; Varghese et al., 2017). Top-down attention allocation is reflected in a reduced N1 component in the frontocentral area (Talsma & Woldorff, 2005). Attentional orienting is indeed sensitive to attentional demand (top-down demand) and robust to the naturalness of the environment.

Table 33

Research implications

Element	Implication
Empirical	Attentional Orienting is sensible to the naturalness of the environment of HCI
	Attentional orienting is sensible to attentional demand (top-down demand) and robust to the naturalness of the environment
Theoretical	The study of attentional mechanisms and their interplay with sensory integration can shed a unique light on psychological-level constructs studied in IS.
Methodological	Perturbation paradigms elicit robust ERPs in ecologically valid HCI tasks
	Multisensory Perturbation techniques, we show that it is possible to generate an attention-orienting neurophysiological response based on task-congruent perturbation
Practical	Attentional Mechanisms are different in a natural task; designing simulators closer to reality elicit more natural cognition

This paper makes a valuable contribution to the fields of cognitive neuroscience and mobile cognition by demonstrating the impact of the interplay between multisensory integration and attention in a quasi-naturalistic environment (Gramann et al., 2011; Ladouce et al., 2016). Moreover, it provides further evidence that exogenous stimulation measures can effectively be employed to examine attention-related constructs within quasi-naturalistic settings. By adopting a unique theoretical and methodological approach, this manuscript explores cognition as an embodied process, encompassing the brain, body, and environment in naturalistic contexts (Matusz et al., 2019; Stangl et al., 2023); thus, it directly contributes to the field of cognitive neuroscience.

This research also contributes to methodological approaches in several ways. First, we show that perturbation paradigms can elicit an ERP that is robust enough during quasi-naturalistic interaction to be measured and sensible to intertwined attentional factors such as top-down forces. It provides evidence that it can effectively measure unconscious attentional mechanisms even without detecting the perturbation. It adds to the NeuroIS methods that can help us understand users during HCI (Dimoka et al., 2012)

Second, we propose a multisensory perturbation technique congruent with the environment that minimizes interference with the HCI task. This technique can generate an attentional-orienting neurophysiological response based on task-congruent perturbation. The potential of this approach is to measure attentional orienting and proxy correlates (e.g., attentional demands) in an unobtrusive manner during naturalistic interaction with the technology. Thus, we posit that this approach is flexible in its application to different settings.

Practitioners can leverage our findings and methodology to evaluate system design. Designing for capturing attentional resources is essential in learning and training environments. Training or performance-oriented simulators use multiple sensory modalities to augment factors associated with learning and transfer performance to the real world (Salas & Cannon-Bowers, 2001). Designers attempt to build digital simulations for training purposes by combining multiple sensory modalities without necessarily evaluating the artifact at the brain functions level. This manuscript shows the utility of neuroscience methods and mental state inference to evaluate technological artifact design choices. Here, we carefully evaluated how aspects of the simulator impact sensory function and cognition.

Measuring the orientation of attention inside those simulators can provide a rich understanding of the neurophysiological effect on users. It also can facilitate the evaluation of design decisions when those technological artifacts are created, and providing methods of evaluation for design principles is a significant component of NeuroIS research (vom Brocke et al., 2020). In this context, measuring the orientation of attention can be valuable to assess if the artifact reaches its designed goals.

4.6.1 Limitations

Our work carries limitations. One notable advantage of the perturbation technique is its ability to minimize interference with the HCI task, rendering it appealing for naturalistic tasks. However, it presents a fundamental limitation regarding the experimental control over the processing of the perturbation through detection and the recording of measurable behavioral responses towards it (e.g., reaction time, perturbation counting). This limitation is well-documented in dual-task paradigms involving perturbation (Kok, 1997). Furthermore, another limitation we encountered relates to the challenge of establishing meaningful connections between neurophysiological responses and natural behaviors within this specific context, primarily due to the complexity inherent in the tasks and the associated cognitive mechanisms (Kirwan et al., 2023).

4.6.2 Future research

First, Further work can be invested into designing natural HCI experiments using multisensory perturbations that are congruent and relevant. For example, task interruption during HCI tasks using audio-visual stimulation can be a naturalistic approach to studying the impact of e-mail interruption and individual performance (Addas & Pinsonneault, 2018). Integrating perturbation techniques congruent to IS tasks could enable the study of natural cognition at work. For example, incorporating visual and auditory notifications from emails or cell phone vibrations can serve as congruent and effective stimuli to evoke physiological responses while working. In such a case, we advise using a similar approach to this thesis by incrementally ensuring the robustness of the neurophysiological responses from laboratory to the real-world.

Secondly, the application of the conceptual framework of sensory integration and oriented attention to relevant IS phenomena is worth considering. The orientation of attention can be linked to psychological constructs such as immersion or presence (Burns & Fairclough, 2015; Marucci et al., 2021). Information System research is often interested in constructs at the individual level that include concepts close to attention and some of its mechanisms. For example, cognitive absorption is defined as "a state of deep involvement with software" and possesses five dimensions: Control, Curiosity, focused immersion, Heightened enjoyment, and temporal dissociation (Agarwal & Karahanna, 2000). Focused immersion refers to "the experience of total engagement where other attentional demands is, in essence, ignored." This dimension is close to the orientation of attention and can be studied at the level of cognitive mechanisms. Our approach study can help us understand the implication of focused immersion during IS use. In the same spirit, our approach can help us investigate the experience of immersion in HCI, particularly real-world dissociation dimension (Jennett et al., 2008). Thus, IS researchers can use this method to explore relevant psychological constructs at the level of cognitive mechanisms.

Thirdly, further work should focus on bridging the gap between user behavior and oriented attention, paving the way for a cohesive research program that unifies cognitive mechanisms with user behaviors. This approach aligns with a recent call for such integration (Kirwan et al., 2023) and holds the potential to enhance our understanding of the complex interplay between cognition and user actions. Moreover, it was a significant limitation of the current work.

4.7 Conclusion

This manuscript introduces an innovative ingenious perturbation technique and provides a conceptual framework for understanding sensory integration in naturalistic settings, specifically focusing on investigating oriented attention in users. This theoretical approach allows us to explore the role of oriented attention and sensory integration within a multisensory environment. Through two studies, we gradually enhance the naturalness of a simulation to ensure that neurophysiological responses to the perturbation processing remain robust in real-world environments. Although the naturalness of the task can influence neurophysiological responses, the results demonstrate that these responses are sensitive to sensory processing and parallel top-down attentional mechanisms. This research contributes to developing theoretical approaches and methodologies for studying cognition in real-world contexts.

References

Addas, S., & Pinsonneault, A. (2015). The many faces of information technology interruptions: a taxonomy and preliminary investigation of their performance effects. *Information Systems Journal*, *25*(3), 231-273.

Addas, S., & Pinsonneault, A. (2018). E-Mail Interruptions and Individual Performance: Is There a Silver Lining? *Management Information Systems Quarterly*, *42*(2), 381-405.

Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS quarterly*, 665-694.

Allison, B. Z., & Polich, J. (2008). Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biological psychology*, *77*(3), 277-283.

Barutchu, A., & Spence, C. (2021). Top–down task-specific determinants of multisensory motor reaction time enhancements and sensory switch costs. *Experimental Brain Research*, *239*(3), 1021-1034. https://doi.org/10.1007/s00221-020-06014-3

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & psychophysics*, *62*(2), 321-332.

Bolton, D. A. E. (2015). The role of the cerebral cortex in postural responses to externally induced perturbations. *Neuroscience and Biobehavioral Reviews*, *57*, 142-155. https://doi.org/10.1016/j.neubiorev.2015.08.014

Burns, C. G., & Fairclough, S. H. (2015). Use of auditory event-related potentials to measure immersion during a computer game. *International Journal of Human-Computer Studies*, *73*, 107-114.

Chen, A., & Karahanna, E. (2018). Life interrupted: The effects of technology-mediated work interruptions on work and nonwork outcomes. *MIS quarterly*, *42*(4), 1023-1042.

Chiel, H. J., & Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in neurosciences*, *20*(12), 553-557.

Choi, I., Lee, J.-Y., & Lee, S.-H. (2018). Bottom-up and top-down modulation of multisensory integration. *Current Opinion in Neurobiology*, *52*, 115-122.

Conrad, C., & Newman, A. (2021). Measuring Mind Wandering During Online Lectures Assessed With EEG [Brief Research Report]. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.697532

Cooper, N., Milella, F., Pinto, C., Cant, I., White, M., & Meyer, G. (2018). The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *Plos One*, *13*(2), e0191846.

Cornelio, P., Velasco, C., & Obrist, M. (2021). Multisensory integration as per technological advances: A review. *Frontiers in neuroscience*, *15*.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics*, *66*(8), 1388-1404.

Dimoka, A., Davis, F. D., Gupta, A., Pavlou, P. A., Banker, R. D., Dennis, A. R., Ischebeck, A., Müller-Putz, G., Benbasat, I., & Gefen, D. (2012). On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS quarterly*, 679-702.

Dimoka, A., Pavlou, P. A., & Davis, F. D. (2011). Research commentary—NeuroIS: The potential of cognitive neuroscience for information systems research. *Information Systems Research*, *22*(4), 687-702.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, *381*(6577), 66-68.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., Polich, J., Reinvang, I., & Van Petten, C. (2009). Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*(11), 1883-1908.

Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature neuroscience*, *8*(12), 1643-1646.

Forschack, N., Nierhaus, T., Müller, M. M., & Villringer, A. (2017). Alpha-band brain oscillations shape the processing of perceptible as well as imperceptible somatosensory stimuli during selective attention. *Journal of neuroscience*, *37*(29), 6983-6994.

Gallagher, M., & Ferrè, E. R. (2018). Cybersickness: a multisensory integration perspective. *Multisensory Research*, *31*(7), 645-674.

Galluch, P. S., Grover, V., & Thatcher, J. B. (2015). Interrupting the workplace: Examining stressors in an information technology context. *Journal of the Association for Information Systems*, *16*(1), 2.

Gaspelin, N., & Luck, S. J. (2018). Top-down" does not mean "voluntary. *Journal of cognition*, *1*(1).

Gottfried, J. A., & Dolan, R. J. (2003). The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception. *Neuron*, *39*(2), 375-386.

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T. P., Lin, C. T., Liao, L. D., & Makeig, S. (2011). Cognition in action: imaging brain/body dynamics in mobile humans. *Rev Neurosci*, *22*(6), 593-608. https://doi.org/10.1515/RNS.2011.047

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., & Parkkonen, L. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in neuroscience*, *7*, 267.

Hagmann, C. E., & Russo, N. (2016). Multisensory integration of redundant trisensory stimulation. *Attention, Perception, & Psychophysics*, *78*(8), 2558-2568.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177-180.

Ho, C., Reed, N., & Spence, C. (2007). Multisensory in-car warning signals for collision avoidance. *Human Factors*, *49*(6), 1107-1114.

Hopfinger, J. B., & West, V. M. (2006). Interactions between endogenous and exogenous attention on cortical visual processing. *Neuroimage*, *31*(2), 774-789.

Horváth, J., Roeber, U., Bendixen, A., & Schröger, E. (2008). Specific or general? The nature of attention set changes triggered by distracting auditory events. *Brain Research*, *1229*, 193-203.

Innes, B. R., & Otto, T. U. (2019). A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent. *Scientific reports*, *9*(1), 2921. https://doi.org/10.1038/s41598-019-39924-6

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, *66*(9), 641-661. https://doi.org/10.1016/j.ijhcs.2008.04.004

Kirwan, C. B., Vance, A., Jenkins, J. L., & Anderson, B. B. (2023). Embracing brain and behaviour: Designing programs of complementary neurophysiological and behavioural studies. *Information Systems Journal*.

Kok, A. (1997). Event-related-potential (ERP) reflections of mental resources: a review and synthesis. *Biological psychology*, *45*(1-3), 19-56.

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2016). Understanding Minds in Real-World Environments: Toward a Mobile Cognition Approach. *Front Hum Neurosci*, *10*, 694. https://doi.org/10.3389/fnhum.2016.00694

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2019). Mobile EEG identifies the re-allocation of attention during real-world activity. *Sci Rep*, *9*(1), 15851. https://doi.org/10.1038/s41598-019-51996-y

Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The curious incident of attention in multisensory integration: bottom-up vs. top-down. *Multisensory Research*, *29*(6-7), 557-583.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177-190.

Marucci, M., Di Flumeri, G., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E. F., Babiloni, F., Betti, V., & Aricò, P. (2021). The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Scientific reports*, *11*(1), 1-15.

Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019). Are we ready for real-world neuroscience? , *31*(3), 327-338.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748.

Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Cognitive Brain Research*, *14*(1), 31-40.

Mulckhuyse, M., & Theeuwes, J. (2010). Unconscious attentional orienting to exogenous cues: A review of the literature. *Acta psychologica*, *134*(3), 299-309.

Müller-Putz, G. R., Riedl, R., & Wriessnegger, S. C. (2015). Electroencephalography (EEG) as a Research Tool in the Information Systems Discipline: Foundations, Measurement, and Applications. *CAIS*, *37*, 46.

Noppeney, U. (2021). Perceptual inference, learning, and attention in a multisensory world. *Annual review of neuroscience*, *44*, 449-473.

Northoff, G. (2018). *The spontaneous brain: from the mind-body to the world-brain problem*.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *Journal of Experimental Psychology: General*, *130*(3), 466.

Polich, J. (1989). Habituation of P300 from auditory stimuli. *Psychobiology*, *17*(1), 19-28.

Quant, S., Adkin, A. L., Staines, W. R., Maki, B. E., & McIlroy, W. E. (2004). The effect of a concurrent cognitive task on cortical potentials evoked by unpredictable balance perturbations. *Bmc Neuroscience*, *5*(1), 1-12.

Quant, S., Maki, B. E., & McIlroy, W. E. (2005). The association between later cortical potentials and later phases of postural reactions evoked by perturbations to upright stance. *Neuroscience letters*, *381*(3), 269-274.

Riedl, R., Davis, F. D., & Hevner, A. R. (2014). Towards a NeuroIS research methodology: intensifying the discussion on methods, tools, and measurement. *Journal of the Association for Information Systems*, *15*(10), I.

Riedl, R., Fischer, T., Léger, P.-M., & Davis, F. D. (2020). A decade of NeuroIS research: progress, challenges, and future directions. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *51*(3), 13-54.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology*, *52*(1), 471-499.

Schröger, E., & Wolff, C. (1998). Attentional orienting and reorienting is indicated by human event-related brain potentials. *Neuroreport*, *9*(15), 3355-3358.

Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual review of physiology*, *25*(1), 545-580.

Stangl, M., Maoz, S. L., & Suthana, N. (2023). Mobile cognition: imaging the human brain in the 'real world'. *Nature reviews neuroscience*, 1-16.

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature reviews neuroscience*, *15*(8), 520-535.

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cerebral cortex*, *17*(3), 679-690.

Talsma, D., Kok, A., Slagter, H. A., & Cipriani, G. (2008). Attentional orienting across the sensory modalities. *Brain and Cognition*, *66*(1), 1-10.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences*, *14*(9), 400-410.

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*(7), 1098-1114.

Tang, X., Wu, J., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience & Biobehavioral Reviews*, *61*, 208-224.

Terkildsen, T., & Makransky, G. (2019). Measuring presence in video games: An investigation of the potential use of physiological measures as indicators of presence. *International Journal of Human-Computer Studies*, *126*, 64-80.

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of experimental psychology: human perception and performance*, *34*(5), 1053.

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2009). Poke and pop: Tactile–visual synchrony increases visual saliency. *Neuroscience letters*, *450*(1), 60-64.
Varghese, J. P., McIlroy, R. E., & Barnett-Cowan, M. (2017). Perturbation-evoked potentials: Significance and application in balance control research. *Neuroscience & Biobehavioral Reviews*, *83*, 267-280.
vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a neurois research agenda with four areas of societal contributions. *European Journal of Information Systems*, *29*(1), 9-24.
Welch, R. B. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In *Advances in psychology* (Vol. 129, pp. 371-387). Elsevier.
Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological bulletin*, *88*(3), 638.
Zhang, X., Zhaoping, L., Zhou, T., & Fang, F. (2012). Neural activities in V1 create a bottom-up saliency map. *Neuron*, *73*(1), 183-192.
Zink, R., Hunyadi, B., Huffel, S. V., & Vos, M. D. (2016). Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks. *J Neural Eng*, *13*(4), 046017. https://doi.org/10.1088/1741-2560/13/4/046017

Conclusion

This thesis offers novel measurement methods and conceptual approaches for studying mental states during human-computer interaction in NeuroIS. Each chapter of the thesis contributes uniquely to this goal. The scoping review in the second chapter revealed current challenges in applied neuroscience within HCI and IS. It was observed that these challenges come, in part, from the limited conceptualization of psychophysiological inference and the slow adoption of state-of-the-art analytical approaches. The subsequent chapters aim to address these limitations through two distinct methodologies. The third chapter employs a data-driven methodology that utilizes state-of-the-art machine learning techniques for mental state decoding. The aim is to generalize the physical brain patterns observed in a synthetic task to a naturalistic setting. The fourth chapter takes a more hypotheticodeductive approach, drawing upon the current understanding of multisensory integration and its complex interaction with attention and various bottom-up and top-down factors. Based on this theoretical foundation, a perturbation-based measure of oriented attention in EEG is developed and tested within naturalistic tasks specific to the IS domain.

The overarching objective of the IT field is to gain a deeper understanding of how the design and utilization of information systems impact users and contribute to the theoretical and practical advancement of the technology (Dimoka et al., 2012; Riedl & Léger, 2016). This thesis aligns with this goal by offering techniques to study cognition during IT usage with high temporal precision, a challenge due to its naturalistic aspect (Riedl et al., 2020). Thus, we directly engage with the call made by Riedl et al. (2014) and vom Brocke et al. (2020) for more research methodology in NeuroIS.

Furthermore, following the recommendations of vom Brocke and Liang (2014), we draw upon knowledge and methodologies from cognitive neuroscience while adhering to the rigorous procedures established within this field. Accordingly, this results in the rigorous benchmarking and validation methodologies applied in training neural networks for mental state decoding, as elaborated in Chapter 3. Additionally, Chapter 4 presents the design of an astute perturbation design and the utilization of advanced data collection architecture and analytical techniques to study oriented attention over time and space in a simulated environment. Our endeavor highlights significant challenges associated with integrating conceptual knowledge, experimental design, and analytical techniques in the study of cognition when considering the complexity and naturalness of the tasks. However, this thesis demonstrates that our rigorous approaches have yielded valuable findings and contributions to NeuroIS.

Our effort illustrates the diverse paths toward improving mental state estimation during interactions, each possessing its strengths and weaknesses. The primary strength of Chapter 3 lies in its end-to-end deep learning approach for decoding mental states, which provides a means to operationalize and measure constructs longitudinally during usage. The output of the machine learning model serves as a measure in a statistical model while preserving the temporal nature of the phenomena, enabling the testing of theories that consider mental states. This approach seamlessly integrates with the current IS approach and offers an alternative method of studying cognition and behaviors. However, it is essential to acknowledge that this technique is still in its early stages and exhibits limitations in terms of generalizability and reliability. Nevertheless, we are optimistic that further research outlined below will enhance its potential as a powerful tool for NeuroIS researchers.

Chapter 4 offers a conceptual foundation for understanding oriented attention in naturalistic contexts, aligning with our objective of investigating cognition as an embodied process encompassing the brain, body, and environment. This theoretical approach enables us to explore oriented attention within the context of sensory integration in a multisensory world. In doing so, we aim to enhance our understanding of cognition during various tasks and achieve a more comprehensive and generalized comprehension of cognitive processes. However, this approach necessitates specific methodological considerations. It required a specific experimental environment and design incorporating comprehensive data collection, including users' EEG, behaviors, and simulated surroundings. One limitation of this inherent complexity of the tasks and the involved cognitive mechanisms is the difficulty in establishing meaningful connections between neurophysiological responses and behaviors within these contexts.

Nevertheless, both empirical studies show the richness of conceptual, methodological, and analytical approaches that can be utilized to enhance state estimation during interactions. These diverse approaches contribute uniquely to the advancement of knowledge and methodology in the field of NeuroIS. Moreover, we are committed to undertaking further research to mitigate this limitation.

This thesis sets the stage for many exciting future research opportunities. Chapter 2, through its scoping review, lays the foundation for further integrating neurophysiological measures into IS artifacts, also referred to as neuro-adaptive systems. This review paves the way for developing a conceptual framework that guides the design and evaluation of neuro-adaptive system artifacts within design science research. Defining the problem, knowledge, and solution spaces for neuro-adaptive systems and their components (e.g., functional components, neuropsychological inferences) is an essential step toward a theoretical base for designing such artifacts, enabling the accumulation of descriptive and prescriptive knowledge in IS. Furthermore, given the inherent naturalness of IS tasks, we have observed that there is still significant work to be done in identifying robust neurophysiological responses that are relevant to our field's cognitive processes. Future research should extend beyond the mere identification of these neurophysiological responses and also focus on the development of evaluation methods to determine the generalizability of such responses to natural IS tasks.

Chapter 3 can inform a future methodology development manuscript for mental state decoding in IS with neurophysiological plausibility assessment building on the methodological framework used. Additionally, our deep-learning approach presents opportunities for further advancement. We identify three critical avenues for improvement. Firstly, implementing relevant data augmentation techniques can address the limitations of small datasets often encountered in experimental studies. Secondly, domain adaptation can be a powerful approach to enhance future deep learning models for mental state adaptation in IS. This technique adapts a model trained on one task to a different yet related task. Domain adaptation can improve mental state estimation generalizability, reliability, and robustness. Thirdly, general model development of mental state estimation (i.e., across subjects) along with fine-tuning techniques show promise in leveraging larger datasets to enhance generalizability and expedite the training process.

In Chapter 4, the focus should shift towards further generalizing the approach to encompass naturalistic IS tasks and enhancing ecological validity. By doing so, we aim to expand the boundaries and generalize our findings to more naturalistic contexts, including real-world IS tasks. Future research should explore the integration of seamless perturbation techniques into computerized tasks that closely resemble real-life scenarios. For example, incorporating visual and auditory notifications from e-mails or cell phone vibrations can serve as congruent and effective stimuli to evoke physiological responses while working. Furthermore, the application of the conceptual framework of sensory integration and oriented attention to various IS phenomena is worth considering. Further work should focus on bridging the gap between user behavior and oriented attention, paving the way for a cohesive research program that unifies cognitive mechanisms with user behaviors. This approach aligns with a recent call for such integration (Kirwan et al., 2023) and holds the potential to enhance our understanding of the complex interplay between cognition and user actions. Finally, it is crucial to recognize that NeuroIS offers a distinctive approach to understanding cognition in naturalistic contexts and provides direct contributions to cognitive neuroscience (Matusz et al., 2019; Stangl et al., 2023). Acknowledging this co-evolutionary perspective, it becomes evident that our ability to extend our understanding to NeuroIS tasks would significantly contribute to our overall comprehension of cognition.

The approaches developed in this thesis, in conjunction with the embodied perspective of cognition, have the potential to enhance the traditional understanding of information systems (IS) phenomena, such as IT usage, interruptions, and technology-mediated tasks. For example, Addas and Pinsonneault (2018) demonstrated that daily incongruent/congruent e-mail interruptions at work impact performance through perceived cognitive workload and mindfulness. Similarly, Chen and Karahanna (2018) discovered that work-related interruptions during leisure time could lead to adverse outcomes, including psychological exhaustion and decreased performance in both work and non-work-related activities. This thesis's methodological and theoretical approaches could complement these phenomena of interest, which focus on psychological aspects. The specific mediating mechanisms concerning users' cognition remain largely unexplored. Therefore, future research in NeuroIS holds promise for significantly advancing our understanding of why and how these detrimental effects occur via cognitive level mechanisms. For example, investigating how attention mechanisms and the integration of sensory information (such as visual and auditory cues) during interruptions impact exhaustion, mindfulness, and performance could provide valuable information. Furthermore, exploring the design of interruptions that alleviate the strain on attentional resources is another important avenue for research.

Future contributions extend beyond the scope of the literature on interruption. For instance, Sullivan et al. (2022) focus on examining the impact of executive functions on IS learning outcomes. This investigation incorporates a three-dimensional conceptualization encompassing working memory, shifting, and inhibition. The findings indicate a positive correlation between these dimensions, declarative knowledge, and post-learning self-efficacy. However, it is essential to note that the measurement of these dimensions relies on behavioral assessments during neurophysiological tasks such as the n-back and the Stroop task. We can enhance the empirical findings by incorporating the approaches developed in this thesis. Chapter 4 of the thesis introduces the concept of oriented attention, which has been shown to be sensitive to executive functions (Talsma et al., 2010). Using this measure can provide valuable insight into the cognitive mechanisms at the physical level during IS learning tasks. Furthermore, it is worth considering that attention mechanisms and neurophysiological responses are influenced by age (Donoghue et al., 2020). Examining age-related cognitive functions can offer a unique perspective in comprehending the challenges older workers and IT users face in the workplace (Tams, 2022).

The approaches developed in this thesis could also potentially contribute to studying contemporary phenomena. Adopting human-centered and cognition-centered approaches could provide valuable insights into the dynamics of human-AI interactions (Rahwan et al., 2019). Assuming that humans and AI will further interact, compete, and cooperate, studying human psychological, behavioral, and cognitive phenomena is essential to understanding how they coexist (Dafoe et al., 2021; Makovi et al., 2023). In IS, multi-agents perspective, or the study of intelligent and human agents, is increasing in popularity (Collins et al., 2021; Padmanabhan et al., 2022). By examining how intelligent machines impact human behavior and cognition, we can better understand the underlying cognitive processes involved in delegation mechanisms when interacting with agentic IS artifacts (Baird & Maruping, 2021). For example, Fügener et al. (2021) show that human-ai collaboration positively affects human performance while negatively impacting unique human knowledge, showing an unexpected negative impact of AI on users. These insights can potentially improve the effectiveness of delegation mechanisms, cooperative behavior, and emergent properties of hybrid collectives. Such results justify the need to incorporate a cognitive level perspective of humans during interactions with AI to refine further interaction mechanisms that better align with human cognitive capabilities. Going beyond performance perspectives, we could study questions such as: What is the symbiotic or parasitic effect on the cognition of humans in hybrid dyads? Does interaction dynamics (e.g., reflexive, supervisory, anticipatory, prescriptive) augment or inhibit specific cognitive mechanisms? What is the impact of generative AI on users' cognition?

I believe that enhancing our ability to investigate cognition within naturalistic tasks is a fundamental aim within the field of NeuroIS. Consequently, it is imperative to critically examine existing methods and develop innovative approaches, as they can significantly advance our understanding of how information systems impact users. This thesis represents a crucial step towards achieving this objective, with the aspiration of enabling further progress and inspiring future research in this direction.

References

Addas, S., & Pinsonneault, A. (2018). E-Mail Interruptions and Individual Performance: Is There a Silver Lining? *MIS quarterly*, *42*(2), 381-405.

Baird, A., & Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS quarterly*, *45*(1).

Chen, A., & Karahanna, E. (2018). Life interrupted: The effects of technology-mediated work interruptions on work and nonwork outcomes. *MIS quarterly*, *42*(4), 1023-1042.

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383.

Dafoe, A., Bachrach, Y., Hadfield, G., Horvitz, E., Larson, K., & Graepel, T. (2021). Cooperative AI: machines must learn to find common ground. *Nature*, *593*(7857), 33-36.

Dimoka, A., Davis, F. D., Gupta, A., Pavlou, P. A., Banker, R. D., Dennis, A. R., Ischebeck, A., Müller-Putz, G., Benbasat, I., & Gefen, D. (2012). On the use of neurophysiological tools in IS research: Developing a research agenda for NeuroIS. *MIS quarterly*, 679-702.

Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A. H., Wallis, J. D., & Knight, R. T. (2020). Parameterizing neural power spectra into periodic and aperiodic components. *Nature neuroscience*, *23*(12), 1655-1665.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *MIS quarterly-Vol*, *45*.

Kirwan, C. B., Vance, A., Jenkins, J. L., & Anderson, B. B. (2023). Embracing brain and behaviour: Designing programs of complementary neurophysiological and behavioural studies. *Information Systems Journal*.

Makovi, K., Sargsyan, A., Li, W., Bonnefon, J.-F., & Rahwan, T. (2023). Trust within human-machine collectives depends on the perceived consensus about cooperative norms. *Nature communications*, *14*(1), 3108.

Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019). Are we ready for real-world neuroscience? , *31*(3), 327-338.

Padmanabhan, B., Sahoo, N., & Burton-Jones, A. (2022). Machine learning in information systems research. *MIS quarterly*, *46*(1), iii-xix.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., & Jackson, M. O. (2019). Machine behaviour. *Nature*, *568*(7753), 477.

Riedl, R., Davis, F. D., & Hevner, A. R. (2014). Towards a NeuroIS research methodology: intensifying the discussion on methods, tools, and measurement. *Journal of the Association for Information Systems*, *15*(10), I.

Riedl, R., Fischer, T., Léger, P.-M., & Davis, F. D. (2020). A decade of NeuroIS research: progress, challenges, and future directions. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *51*(3), 13-54.

Riedl, R., & Léger, P.-M. (2016). Fundamentals of NeuroIS. *Studies in Neuroscience, Psychology and Behavioral Economics. Springer, Berlin, Heidelberg*.

Stangl, M., Maoz, S. L., & Suthana, N. (2023). Mobile cognition: imaging the human brain in the 'real world'. *Nature reviews neuroscience*, 1-16.

Sullivan, Y. W., Davis, F. D., & Koh, C. E. (2022). Executive functions and information systems learning. *MIS quarterly*, *46*(2).

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences*, *14*(9), 400-410.

Tams, S. (2022). Helping older workers realize their full organizational potential: a moderated mediation model of age and it-enabled task performance. *MIS quarterly*, *46*(1).

vom Brocke, J., Hevner, A., Léger, P. M., Walla, P., & Riedl, R. (2020). Advancing a neurois research agenda with four areas of societal contributions. *European Journal of Information Systems*, *29*(1), 9-24.

vom Brocke, J., & Liang, T.-P. (2014). Guidelines for Neuroscience Studies in Information Systems Research. *Journal of management information systems*, *30*(4), 211-234. https://doi.org/10.2753/MIS0742-1222300408

Appendix

A1 Chapter 2

A1.1 PRISMA checklist

Table 34

Preferred reporting items for systematic reviews and meta-analyses extension for scoping reviews (prisma-scr) checklist (Tricco et al., 2018)

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
TITLE			
Title	1	Identify the report as a scoping review.	Page 1
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, conclusions and keywords that relate to the review questions and objectives.	Page 1
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	Pages 1-3 and 9
Objectives	4	Provide an explicit statement of the review question and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	Page 9
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	Unregistered review
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	Page 13
Information	7	Describe all information sources in the search	Page 12

sources*		(e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	Page 11 and Appendix
Selection of sources of evidence†	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	Page 13
Data charting process‡	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	Page 14
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	Appendix
Critical appraisal of individual sources of evidence§	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	Not conducted
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	Page 38
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	Pages 15-16
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	Pages 15-16
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	Not conducted
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	Results section
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	Results tables

DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	Discussion
Limitations	20	Discuss the limitations of the scoping review process.	Discussion
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	Discussion
FUNDING			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	Ivado, FRQNT

Notes. Relative paging in the scoping review.

A1.2 Scoping process

A1.2.1 Search queries

Table 35

Phase 1 – Tuning queries

Database	# of results	Query	Comment
Web of Science	215	TS=(("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "aerospace" OR "aero*" OR "aeronautics" OR "automotive" OR "socio-technic*") AND ("Cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition")	Observation: A lot of geriatric literature due to aerobic and mental states
Web of Science	25	TS=(("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "aerospace" OR "aero*" OR "aeronautics" OR "automotive" OR "socio-technic*") AND ("Cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition") NOT ("aerobic*" OR “aerodynamic” OR “aerosol” OR “aerogenes”))	Observation: Search operator NOT to exclude “aerobic, aerodynamic, aerosol” cleans the results enough. We increase the breath of the query.
Web of Science	1059	TS=(("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "aerospace" OR "aero*" OR "aeronautics" OR "transport*" OR "automotive" OR "socio-technic*" OR "automat*" OR "digital automat*") AND ("cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition") NOT ("aerobic*" OR “aerodynamic” OR “aerosol” OR “aerogenes”))	Observation: geriatric, psychiatry, and neurology papers We increase the NOT

Web of Science	787	TS=(("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "aerospace" OR "aero*" OR "aeronautics" OR "transport*" OR "automotive" OR "socio-technic*" OR "automat*" OR "digital automat*") AND ("cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition") NOT ("aerobic*" OR “aerodynamic” OR “aerosol” OR “aerogenes”)) NOT WC=("Geriatrics Gerontology" OR “Psychiatry” OR “Clinical Neurology”)	

Table 36

Phase 2 – Extending queries

WoS 2nd pass (02-28)	4210	TS=(("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "aerospace" OR "aero*" OR "aeronautics" OR "transport*" OR "automotive" OR "socio-technic*" OR "automat*" OR "digital automat*" OR "cloud computing" OR "cognitive computing" OR "enterprise systems" OR "information system") AND ("cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition" OR "cognitive function" OR "cognitive mental state" OR "implicit cognitive processes" OR "psychological state" OR "cognition" OR "emotion") AND ("adaptive" OR "adapt*" OR "adaptive automation" OR "adaptive systems" OR "assistive" OR "inclusive design" OR "support" OR "human-systems" OR "human-systems inclusion" OR "human-systems integration" OR "human-machine systems" OR "control systems" OR "Dependable systems" OR "Dependability" OR "rehabilitation" OR "augmented cognition" OR "real-time" OR "learner modeling" OR "human-automation performance" OR "cognitive state profile" OR "mitigation strategies" OR "biofeedback" OR "Real-time adaptive system" OR "adaptable automation" OR "dynamic function allocation" OR "fallback" OR "Human robot interaction" OR "Robot assisted" OR "psychological adjusting" OR "biofeedback" OR "cognitive performance enhancement" OR "Adaptive assistance" OR "Human-autonomy-teaming" OR "mental support" OR "human machine interaction" OR "dynamic adaptation" OR "safety, Supervisory control" OR "Dual control" OR "adaptive human-automation systems"))	131 papers screened
Scopus	10,199	TITLE-ABS-KEY (("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "aerospace" OR "aero*" OR "aeronautics" OR "transport*" OR "automotive" OR "socio-technic*" OR "automat*" OR "digital automat*" OR "cloud computing" OR "cognitive computing" OR "enterprise systems" OR "information system") AND ("cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition" OR "cognitive function" OR "cognitive mental state" OR "implicit cognitive processes" OR "psychological state" OR "cognition" OR "emotion") AND ("adaptive" OR "adapt*" OR "adaptive automation" OR "adaptive systems" OR "assistive" OR "inclusive design" OR "support" OR "human-systems" OR "human-systems inclusion" OR "human-systems integration" OR "human-machine systems" OR "control systems" OR "Dependable systems" OR "Dependability" OR "rehabilitation" OR "augmented cognition" OR "real-time" OR "learner modeling" OR "human-automation performance" OR "cognitive state profile" OR	185 papers screened for new keywords

		"mitigation strategies" OR "biofeedback" OR "Real-time adaptive system" OR "adaptable automation" OR "dynamic function allocation" OR "fallback" OR "Human robot interaction" OR "Robot assisted" OR "psychological adjusting" OR "biofeedback" OR "cognitive performance enhancement" OR "Adaptive assistance" OR "Human-autonomy-teaming" OR "mental support" OR "human machine interaction" OR "dynamic adaptation" OR "safety, Supervisory control" OR "Dual control" OR "adaptive human-automation systems"))	
PubMed	1225	(("manifactur*"[tiab] OR "smart factory"[tiab] OR "industry 4.0"[tiab] OR "aerospace"[tiab] OR "aero*"[tiab] OR "aeronautics"[tiab] OR "transport*"[tiab] OR "automotive"[tiab] OR "socio technic*"[tiab] OR "automat*"[tiab] OR "digital automat*"[tiab] OR "cloud computing"[tiab] OR "cognitive computing"[tiab] OR "enterprise systems"[tiab] OR "information system"[tiab]) AND ("cognitive state"[tiab] OR "mental state"[tiab] OR "psychological state"[tiab] OR "mental process"[tiab] OR "mental condition"[tiab] OR "cognitive function"[tiab] OR "cognitive mental state"[tiab] OR "implicit cognitive processes"[tiab] OR "psychological state"[tiab] OR "cognition"[tiab] OR "emotion"[tiab]) AND ("adaptive"[tiab] OR "adapt*"[tiab] OR "adaptive automation"[tiab] OR "adaptive systems"[tiab] OR "assistive"[tiab] OR "inclusive design"[tiab] OR "support"[tiab] OR "human-systems"[tiab] OR "human-systems inclusion"[tiab] OR "human-systems integration"[tiab] OR "human-machine systems"[tiab] OR "control systems"[tiab] OR "Dependable systems"[tiab] OR "Dependability"[tiab] OR "augmented cognition"[tiab] OR "real-time"[tiab] OR "learner modeling"[tiab] OR "human-automation performance"[tiab] OR "mitigation strategies"[tiab] OR "biofeedback"[tiab] OR "adaptable automation"[tiab] OR "fallback"[tiab] OR "Human robot interaction"[tiab] OR "Robot assisted"[All Fields] OR "biofeedback"[All Fields] OR "cognitive performance enhancement"[tiab] OR "Adaptive assistance"[tiab] OR "Human-autonomy-teaming"[tiab] OR "mental support"[tiab] OR "human machine interaction"[tiab] OR "dynamic adaptation"[tiab] OR "Dual control"[tiab] OR "adaptive human-automation systems"[tiab]))	158 papers reviewed
ACM (02/28)	3009	[[Full Text: "manifactur*"] OR [Full Text: "smart manifactur*"] OR [Full Text: "smart factory"] OR [Full Text: "connected manufacturing"] OR [Full Text: "industry 4.0"] OR [Full Text: "aerospace"] OR [Full Text: "aero*"] OR [Full Text: "aeronautics"] OR [Full Text: "transport*"] OR [Full Text: "automotive"] OR [Full Text: "socio-technic*"] OR [Full Text: "automat*"] OR [Full Text: "digital automat*"] OR [Full Text: "cloud computing"] OR [Full Text: "cognitive computing"] OR [Full Text: "enterprise systems"] OR [Full Text: "information system"]] AND [[Full Text: "cognitive state"] OR [Full Text: "mental state"] OR [Full Text: "psychological state"] OR [Full Text: "mental process"] OR [Full Text: "mental condition"] OR [Full Text:	Book chapter were not considered (only conference and journal

		"cognitive function"] OR [Full Text: "cognitive mental state"] OR [Full Text: "implicit cognitive processes"] OR [Full Text: "psychological state"] OR [Full Text: "cognition"] OR [Full Text: "emotion"]] AND [[Full Text: "adaptive"] OR [Full Text: "adapt*"] OR [Full Text: "adaptive automation"] OR [Full Text: "adaptive systems"] OR [Full Text: "assistive"] OR [Full Text: "inclusive design"] OR [Full Text: "support"] OR [Full Text: "human-systems"] OR [Full Text: "human-systems inclusion"] OR [Full Text: "human-systems integration"] OR [Full Text: "human-machine systems"] OR [Full Text: "control systems"] OR [Full Text: "dependable systems"] OR [Full Text: "dependability"] OR [Full Text: "rehabilitation"] OR [Full Text: "augmented cognition"] OR [Full Text: "real-time"] OR [Full Text: "learner modeling"] OR [Full Text: "human-automation performance"] OR [Full Text: "cognitive state profile"] OR [Full Text: "mitigation strategies"] OR [Full Text: "biofeedback"] OR [Full Text: "real-time adaptive system"] OR [Full Text: "adaptable automation"] OR [Full Text: "dynamic function allocation"] OR [Full Text: "fallback"] OR [Full Text: "human robot interaction"] OR [Full Text: "robot assisted"] OR [Full Text: "psychological adjusting"] OR [Full Text: "biofeedback"] OR [Full Text: "cognitive performance enhancement"] OR [Full Text: "adaptive assistance"] OR [Full Text: "human-autonomy-teaming"] OR [Full Text: "mental support"] OR [Full Text: "human machine interaction"] OR [Full Text: "dynamic adaptation"] OR [Full Text: "safety, supervisory control"] OR [Full Text: "dual control"] OR [Full Text: "adaptive human-automation systems"]]	manuscripts) due to the absence of abstract and keywords when extracted from ACM database platform. 115 papers screened

Table 37

Phase 3 – Extending queries

WoS 3rd pass	22 287	TS=("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "aerospace" OR "aero*" OR "aeronautics" OR "transport*" OR "automotive" OR "Automat*" OR "digital automat*" OR "cloud computing" OR "cognitive computing" OR "enterprise systems" OR "information system" OR "robotics" OR "distance education" OR "medical" OR "medication service" OR "medical" OR "autonomous vehicules" OR "cyber-physical systems" OR "cognitive systems engineering" OR "digital assistance systems" OR "adaptive instructional systems" OR "agent based systems" OR "Cognitive Medical Robots" OR "Smart Environment" OR "personalized medicine" OR "Ambient Intelligence" OR "smart health environment" OR "industrial robots" OR "autonomous operations" OR "medical robots and systems" OR "smart cockpit" OR "digital twin" OR "one-to-many systems") AND (TS=("Cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition" OR "cognitive function" OR "cognitive mental state" OR " implicit cognitive processes" OR "psychological state" OR "cognition" OR "emotion" OR "engagement" OR "workload" OR "mental workload" OR "Situational awareness" OR "Multitasking" OR "attention" OR "drowsiness" OR "distraction" OR "alertness" OR "fatigue" OR "Boredom" OR "Anxiety" OR "stress" OR "emotion" OR "Vigilance" OR "Working memory" OR "intent" OR "distraction" OR "alertness" OR "confusion" OR "human intention" OR "cognitive absorption" OR "mental diseases" OR "information overload" OR "cognitive readiness") AND TS=("adaptive" OR "adaptive automation" OR "adaptive systems" OR "assistive" OR "inclusive design" OR "human-systems" OR "human-systems inclusion" OR "human-systems integration" OR "human-machine systems" OR "control systems" OR "Dependable systems" OR "Dependability" OR "augmented cognition" OR "real-time " OR "learner modeling" OR "human-automation performance" OR "cognitive state profile" OR "mitigation strategies" OR "biofeedback" OR "Real-time adaptive system" OR "adaptable automation" OR "dynamic function allocation" OR "fallback" OR "Human robot interaction" OR "Robot assisted" OR "psychological adjusting" OR "biofeedback" OR "cognitive performance enhancement" OR "Adaptive assistance" OR "Human-autonomy-teaming" OR "mental support" OR "human machine interaction" OR "dynamic adaptation" OR "Supervisory control" OR "Dual control" OR "adaptive human-automation systems" OR "grasp planning" OR "collaborative robots" OR "cobots" OR "admittance control" OR "computer-aided diagnosis" OR "social robots" OR "Robot Assisted Training" OR "Assistive Robotics" OR "supervisory control" OR

		”human in-the-loop” OR ”Cognitive assistant” OR ”intelligent assistant” OR ”virtual assistant” OR ”virtual agent” OR ”adaptive interface” OR ”neurofeedback” OR ”driver analyzer” OR ”driver model” OR ”cognitive automation” OR ”intuitive cognition” OR ”embodied cognition” OR ”adaptive learning” OR ”cognitive monitoring”))
Scopus	58 229	(TITLE-ABS-KEY ("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "aerospace" OR "aero*" OR "aeronautics" OR "transport*" OR "automotive" OR "Automat*" OR "digital automat*" OR "cloud computing" OR "cognitive computing" OR "enterprise systems" OR "information system" OR "robotics" OR "distance education" OR "medical" OR "medication service" OR "medical" OR "autonomous vehicules" OR "cyber-physical systems" OR "cognitive systems engineering" OR "digital assistance systems" OR "adaptive instructional systems" OR "agent based systems" OR "Cognitive Medical Robots" OR "Smart Environment" OR "personalized medicine" OR "Ambient Intelligence" OR "smart health environment" OR "industrial robots" OR "autonomous operations" OR "medical robots and systems" OR "smart cockpit" OR "digital twin" OR "one-to-many systems")) AND (TITLE-ABS-KEY ("Cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition" OR "cognitive function" OR "cognitive mental state" OR " implicit cognitive processes" OR "psychological state" OR "cognition" OR "emotion" OR "engagement" OR "workload" OR "mental workload" OR "Situational awareness" OR "Multitasking" OR "attention" OR "drowsiness" OR "distraction" OR "alertness" OR "fatigue" OR "Boredom" OR "Anxiety" OR "stress" OR "emotion" OR "Vigilance" OR "Working memory" OR "intent" OR "distraction" OR "alertness" OR "confusion" OR "human intention" OR "cognitive absorption" OR "mental diseases" OR "information overload" OR "cognitive readiness") AND (TITLE-ABS-KEY("adaptive" OR "adaptive automation" OR "adaptive systems" OR "assistive" OR "inclusive design" OR "human-systems" OR "human-systems inclusion" OR "human-systems integration" OR "human-machine systems" OR "control systems" OR "Dependable systems" OR "Dependability" OR "augmented cognition" OR "real-time " OR "learner modeling" OR "human-automation performance" OR "cognitive state profile" OR "mitigation strategies" OR "biofeedback" OR "Real-time adaptive system" OR "adaptable automation" OR "dynamic function allocation" OR "fallback" OR "Human robot interaction" OR "Robot assisted" OR "psychological adjusting" OR "biofeedback" OR "cognitive performance enhancement" OR "Adaptive assistance" OR "Human-autonomy-teaming" OR "mental support" OR "human machine interaction" OR "dynamic adaptation" OR "Supervisory control" OR "Dual control" OR "adaptive human-automation

		systems" OR "grasp planning" OR "collaborative robots" OR "cobots" OR "admittance control" OR "computer-aided diagnosis" OR "social robots" OR "Robot Assisted Training" OR "Assistive Robotics" OR "supervisory control" OR "human in-the-loop" OR "Cognitive assistant" OR "intelligent assistant" OR "virtual assistant" OR "virtual agent" OR "adaptive interface" OR "neurofeedback" OR "driver analyzer" OR "driver model" OR "cognitive automation" OR "intuitive cognition" OR "embodied cognition" OR "adaptive learning" OR "cognitive monitoring")))
PubMed	7 174	("manifactur*"[tiab] OR "smart manifactur*"[tiab] OR "smart factory"[tiab] OR "connected manufacturing"[tiab] OR "industry 4.0"[tiab] OR "aerospace"[tiab] OR "aero*"[tiab] OR "aeronautics"[tiab] OR "transport*"[tiab] OR "automotive"[tiab] OR "Automat*"[tiab] OR "digital automat*"[tiab] OR "cloud computing"[tiab] OR "cognitive computing"[tiab] OR "enterprise systems"[tiab] OR "information system"[tiab] OR "robotics"[tiab] OR "distance education"[tiab] OR "medical"[tiab] OR "medication service"[tiab] OR "medical"[tiab] OR "autonomous vehicules" OR "cyber-physical systems"[tiab] OR "cognitive systems engineering"[tiab] OR "digital assistance systems"[tiab] OR "adaptive instructional systems"[tiab] OR "agent based systems"[tiab] OR "Cognitive Medical Robots"[tiab] OR "Smart Environment"[tiab] OR "personalized medicine"[tiab] OR "Ambient Intelligence"[tiab] OR "smart health environment"[tiab] OR "industrial robots"[tiab] OR "autonomous operations"[tiab] OR "medical robots and systems"[tiab] OR "smart cockpit"[tiab] OR "digital twin"[tiab] OR "one-to-many systems"[tiab]) AND (("Cognitive state"[tiab] OR "mental state"[tiab] OR "psychological state"[tiab] OR "mental process"[tiab] OR "mental condition"[tiab] OR "cognitive function"[tiab] OR "cognitive mental state"[tiab] OR " implicit cognitive processes"[tiab] OR "psychological state"[tiab] OR "cognition"[tiab] OR "emotion"[tiab] OR "engagement"[tiab] OR "workload"[tiab] OR "mental workload"[tiab] OR "Situational awareness"[tiab] OR "Multitasking"[tiab] OR "attention"[tiab] OR "drowsiness"[tiab] OR "distraction"[tiab] OR "alertness"[tiab] OR "fatigue"[tiab] OR "Boredom"[tiab] OR "Anxiety"[tiab] OR "stress"[tiab] OR "emotion"[tiab] OR "Vigilance"[tiab] OR "Working memory"[tiab] OR "intent"[tiab] OR "distraction"[tiab] OR "alertness"[tiab] OR "confusion"[tiab] OR "human intention"[tiab] OR "cognitive absorption"[tiab] OR "mental diseases"[tiab] OR "information overload"[tiab] OR "cognitive readiness"[tiab]) AND (("adaptive"[tiab] OR "adaptive automation"[tiab] OR "adaptive systems"[tiab] OR "assistive"[tiab] OR "inclusive design"[tiab] OR "human-systems"[tiab] OR "human-systems inclusion"[tiab] OR "human-systems integration"[tiab] OR "human-machine systems"[tiab] OR "control systems"[tiab] OR "Dependable systems"[tiab] OR "Dependability"[tiab] OR "augmented cognition"[tiab] OR "real-time"[tiab] OR "learner modeling"[tiab] OR "human-automation performance"[tiab]

		OR "cognitive state profile"[tiab] OR "mitigation strategies"[tiab] OR "biofeedback"[tiab] OR "Real-time adaptive system"[tiab] OR "adaptable automation"[tiab] OR "dynamic function allocation"[tiab] OR "fallback"[tiab] OR "Human robot interaction"[tiab] OR "Robot assisted"[tiab] OR "psychological adjusting"[tiab] OR "biofeedback"[tiab] OR "cognitive performance enhancement"[tiab] OR "Adaptive assistance"[tiab] OR "Human-autonomy-teaming"[tiab] OR "mental support"[tiab] OR "human machine interaction"[tiab] OR "dynamic adaptation"[tiab] OR "Supervisory control"[tiab] OR "Dual control"[tiab] OR "adaptive human-automation systems"[tiab] OR "grasp planning"[tiab] OR "collaborative robots"[tiab] OR "cobots"[tiab] OR "admittance control"[tiab] OR "computer-aided diagnosis"[tiab] OR "social robots"[tiab] OR "Robot Assisted Training"[tiab] OR "Assistive Robotics"[tiab] OR "supervisory control"[tiab] OR "human in-the-loop"[tiab] OR "Cognitive assistant"[tiab] OR "intelligent assistant"[tiab] OR "virtual assistant"[tiab] OR "virtual agent"[tiab] OR "adaptive interface"[tiab] OR "neurofeedback"[tiab] OR "driver analyzer"[tiab] OR "driver model"[tiab] OR "cognitive automation"[tiab] OR "intuitive cognition"[tiab] OR "embodied cognition"[tiab] OR "adaptive learning"[tiab] OR "cognitive monitoring"[tiab])))
ACM	35853	[[Full Text: "manifactur*"] OR [Full Text: "smart manifactur*"] OR [Full Text: "smart factory"] OR [Full Text: "connected manufacturing"] OR [Full Text: "industry 4.0"] OR [Full Text: "aerospace"] OR [Full Text: "aero*"] OR [Full Text: "aeronautics"] OR [Full Text: "transport*"] OR [Full Text: "automotive"] OR [Full Text: "automat*"] OR [Full Text: "digital automat*"] OR [Full Text: "cloud computing"] OR [Full Text: "cognitive computing"] OR [Full Text: "enterprise systems"] OR [Full Text: "information system"] OR [Full Text: "robotics"] OR [Full Text: "distance education"] OR [Full Text: "medical"] OR [Full Text: "medication service"] OR [Full Text: "medical"] OR [Full Text: "autonomous vehicules"] OR [Full Text: "cyber-physical systems"] OR [Full Text: "cognitive systems engineering"] OR [Full Text: "digital assistance systems"] OR [Full Text: "adaptive instructional systems"] OR [Full Text: "agent based systems"] OR [Full Text: "cognitive medical robots"] OR [Full Text: "smart environment"] OR [Full Text: "personalized medicine"] OR [Full Text: "ambient intelligence"] OR [Full Text: "smart health environment"] OR [Full Text: "industrial robots"] OR [Full Text: "autonomous operations"] OR [Full Text: "medical robots and systems"] OR [Full Text: "smart cockpit"] OR [Full Text: "digital twin"] OR [Full Text: "one-to-many systems"]] AND [[Full Text: "cognitive state"] OR [Full Text: "mental state"] OR [Full Text: "psychological state"] OR [Full Text: "mental process"] OR [Full Text: "mental condition"] OR [Full Text: "cognitive function"] OR [Full Text: "cognitive mental state"] OR [Full Text: " implicit cognitive processes"] OR [Full Text: "psychological state"] OR [Full Text: "cognition"] OR [Full Text: "emotion"] OR [Full Text: "engagement"] OR [Full Text: "workload"] OR [Full Text: "mental workload"] OR [Full Text: "situational awareness"] OR [Full Text: "multitasking"] OR [Full Text: "attention"] OR [Full Text: "drowsiness"] OR [Full Text: "distraction"] OR [Full Text: "alertness"] OR [Full Text:

		"fatigue"] OR [Full Text: "boredom"] OR [Full Text: "anxiety"] OR [Full Text: "stress"] OR [Full Text: "emotion"] OR [Full Text: "vigilance"] OR [Full Text: "working memory"] OR [Full Text: "intent"] OR [Full Text: "distraction"] OR [Full Text: "alertness"] OR [Full Text: "confusion"] OR [Full Text: "human intention"] OR [Full Text: "cognitive absorption"] OR [Full Text: "mental diseases"] OR [Full Text: "information overload"] OR [Full Text: "cognitive readiness"]] AND [[All: "adaptive"] OR [All: "adaptive automation"] OR [All: "adaptive systems"] OR [All: "assistive"] OR [All: "inclusive design"] OR [All: "human-systems"] OR [All: "human-systems inclusion"] OR [All: "human-systems integration"] OR [All: "human-machine systems"] OR [All: "control systems"] OR [All: "dependable systems"] OR [All: "dependability"] OR [All: "augmented cognition"] OR [All: "real-time "] OR [All: "learner modeling"] OR [All: "human-automation performance"] OR [All: "cognitive state profile"] OR [All: "mitigation strategies"] OR [All: "biofeedback"] OR [All: "real-time adaptive system"] OR [All: "adaptable automation"] OR [All: "dynamic function allocation"] OR [All: "fallback"] OR [All: "human robot interaction"] OR [All: "robot assisted"] OR [All: "psychological adjusting"] OR [All: "biofeedback"] OR [All: "cognitive performance enhancement"] OR [All: "adaptive assistance"] OR [All: "human-autonomy-teaming"] OR [All: "mental support"] OR [All: "human machine interaction"] OR [All: "dynamic adaptation"] OR [All: "supervisory control"] OR [All: "dual control"] OR [All: "adaptive human-automation systems"] OR [All: "grasp planning"] OR [All: "collaborative robots"] OR [All: "cobots"] OR [All: "admittance control"] OR [All: "computer-aided diagnosis"] OR [All: "social robots"] OR [All: "robot assisted training"] OR [All: "assistive robotics"] OR [All: "supervisory control"] OR [All: "human in-the-loop"] OR [All: "cognitive assistant"] OR [All: "intelligent assistant"] OR [All: "virtual assistant"] OR [All: "virtual agent"] OR [All: "adaptive interface"] OR [All: "neurofeedback"] OR [All: "driver analyzer"] OR [All: "driver model"] OR [All: "cognitive automation"] OR [All: "intuitive cognition"] OR [All: "embodied cognition"] OR [All: "adaptive learning"] OR [All: "cognitive monitoring"]]

Table 38

Phase 4 – Validation queries

WoS	11700	TS=("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "smart environment" OR "digital twin" OR "internet of things" OR "control system" OR "dependable system" OR "supervisory control" OR "dual control" OR "aerospace" OR "aero*" OR "aeronautics" OR "transport*" OR "automotive" OR "autonomous vehicle" OR "smart cockpit" OR "air traffic control" OR "autonomous operation" OR "automated driving" OR "smart autonomous vehicle system" OR "automat*" OR "digital automat*" OR "digital assistance system" OR "automated decision aid" OR " cloud computing" OR "cognitive computing" OR "enterprise system" OR "information system" OR "cognitive systems engineering" OR "agent based system" OR "ambient intelligence" OR "one-to-many system" OR "human-systems" OR "human-machine system" OR "human-autonomy-teaming" OR "cognitive assistant" OR "intelligent assistant" OR "virtual assistant" OR "virtual agent" OR "synthetic teammate" OR "intelligent human-machine interaction" OR "cognitive assistance system" OR "emotional-based agent" OR "robotics" OR "cyber-physical system" OR "industrial robot" OR "social robot" OR "evolutionary robotics" OR "cognitive robotics" OR "aerial robotic" OR "teleoperation" OR "telerobotics" OR "cyber-physical-human-system" OR "human-cyber-physical system" OR "human robot interaction" OR "human machine interaction" OR "collaborative robot" OR "cobot" OR "physical human-robot interaction" OR "physical-robot-human interaction" OR "shared robotic task" OR "human-robot team" OR "closed-loop human-robot interaction" OR "real-time human-robot interaction" OR "robotic symbiotic network" OR "human-robot collaboration" OR "safe physical human–robot collaboration" OR "brain mediated human-robot interaction" OR "admittance control" OR "distance education" OR "cognitive medical robot" OR "smart health environment" OR "medical robots and system" OR "robot assisted surgery" OR "robotic surgical procedures" OR "computer-aided diagnosis" OR "neuronavigation") AND (TS=("cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition" OR "cognitive function" OR "cognitive mental state" OR "cognitive processes" OR "psychological state" OR "cognition" OR "emotion" OR "engagement" OR "workload" OR "mental workload" OR "situational awareness" OR "multitasking" OR "attention" OR "drowsiness" OR "distraction" OR "alertness" OR "fatigue" OR "boredom" OR "anxiety" OR "stress" OR "emotion" OR "vigilance" OR "working memory" OR "intent" OR "distraction" OR "alertness" OR "confusion" OR "human intention" OR "cognitive absorption" OR "information overload" OR "cognitive readiness" OR "sleepiness" OR "attentional tunneling" OR "vigilance" OR "cognitive workload" OR "inattention") AND TS=("adaptive" OR "adapt*" OR "adaptive automation" OR "adaptive systems" OR "assistive" OR "inclusive design" OR "human-automation performance" OR "real-time adaptive system" OR "adaptable automation" OR "dynamic function allocation" OR "fallback" OR "adaptive assistance" OR "dynamic adaptation" OR "adaptive human-automation systems" OR "human in-the-loop" OR "adaptive interface" OR "cognitive automation" OR

		“adaptive learning” OR “adaptive control” OR “system adaptation” OR “adaptive workload allocation” OR “adaptive aiding” OR “adaptive cruise control” OR “flexible automation” OR “adaptive mitigation strategies” OR “human-systems inclusion” OR “human-systems integration” OR “multimodal interaction” OR “biofeedback” OR “augmented cognition” OR “psychological adjusting” OR “biofeedback” OR “cognitive performance enhancement” OR “cognitive enhancement” OR “mental support” OR “neurofeedback” OR “cognitive monitoring”)) NOT TS=(“Plant” OR “Rehabilitation” OR “Disease” OR “Therapy” OR “Therapeutic” OR “Microbiology” OR “Microbial” OR “Pathogen” OR “Aerobic” OR “aerodynamic” OR “aerosol” OR “aerogenes”)
Scopus	16792	((TITLE-ABS-KEY ("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "smart environment" OR "digital twin" OR "internet of things" OR "control system" OR "dependable system" OR "supervisory control" OR "dual control" OR "aerospace" OR "aero*" OR "aeronautics" OR "transport*" OR "automotive" OR "autonomous vehicle" OR "smart cockpit" OR "air traffic control" OR "autonomous operation" OR "automated driving" OR "smart autonomous vehicle system" OR "automat*" OR "digital automat*" OR "digital assistance system" OR "automated decision aid" OR " cloud computing" OR "cognitive computing" OR "enterprise system" OR "information system" OR "cognitive systems engineering" OR "agent based system" OR "ambient intelligence" OR "one-to-many system" OR "human-systems" OR "human-machine system" OR "human-autonomy-teaming" OR "cognitive assistant" OR "intelligent assistant" OR "virtual assistant" OR "virtual agent" OR "synthetic teammate" OR "intelligent human-machine interaction" OR "cognitive assistance system" OR "emotional-based agent" OR "robotics" OR "cyber-physical system" OR "industrial robot" OR "social robot" OR "evolutionary robotics" OR "cognitive robotics" OR "aerial robotic" OR "teleoperation" OR "telerobotics" OR "cyber-physical-human-system" OR "human-cyber-physical system" OR "human robot interaction" OR "human machine interaction" OR "collaborative robot" OR "cobot" OR "physical human-robot interaction" OR "physical-robot-human interaction" OR "shared robotic task" OR "human-robot team" OR "closed-loop human-robot interaction" OR "real-time human-robot interaction" OR "robotic symbiotic network" OR "human-robot collaboration" OR "safe physical human--robot collaboration" OR "brain mediated human-robot interaction" OR "admittance control" OR "distance education" OR "cognitive medical robot" OR "smart health environment" OR "medical robots and system" OR "robot assisted surgery" OR "robotic surgical procedures" OR "computer-aided diagnosis" OR "neuronavigation")) AND (TITLE-ABS-KEY ("cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition" OR "cognitive function" OR "cognitive mental state" OR "cognitive processes" OR "psychological state" OR "cognition" OR "emotion" OR "engagement" OR "workload" OR "mental workload" OR "situational awareness" OR "multitasking" OR "attention" OR "drowsiness" OR "distraction" OR "alertness" OR "fatigue" OR "boredom" OR "anxiety" OR "stress" OR "emotion" OR "vigilance" OR "working memory" OR "intent" OR "distraction" OR "alertness" OR "confusion" OR "human intention" OR "cognitive absorption" OR "information overload" OR "cognitive readiness" OR "sleepiness" OR "attentional tunneling" OR "vigilance"

		OR "cognitive workload" OR "inattention")) AND (TITLE-ABS-KEY ("adaptive" OR "adapt*" OR "adaptive automation" OR "adaptive systems" OR "assistive" OR "inclusive design" OR "human-automation performance" OR "real-time adaptive system" OR "adaptable automation" OR "dynamic function allocation" OR "fallback" OR "adaptive assistance" OR "dynamic adaptation" OR "adaptive human-automation systems" OR "human in-the-loop" OR "adaptive interface" OR "cognitive automation" OR "adaptive learning" OR "adaptive control" OR "system adaptation" OR "adaptive workload allocation" OR "adaptive aiding" OR "adaptive cruise control" OR "flexible automation" OR "adaptive mitigation strategies" OR "human-systems inclusion" OR "human-systems integration" OR "multimodal interaction" OR "biofeedback" OR "augmented cognition" OR "psychological adjusting" OR "biofeedback" OR "cognitive performance enhancement" OR "cognitive enhancement" OR "mental support" OR "neurofeedback" OR "cognitive monitoring")) AND NOT (TITLE-ABS-KEY ("Plant" OR "Rehabilitation" OR "Disease" OR "Therapy" OR "Therapeutic" OR "Microbiology" OR "Microbial" OR "Pathogen" OR "Aerobic" OR "aerodynamic" OR "aerosol" OR "aerogenes"))) AND PUBYEAR > 2011 AND (EXCLUDE (SUBJAREA,"BIOC"))
PubMed	2270	((“manifactur*”[tiab] OR “smart manifactur*”[tiab] OR “smart factory”[tiab] OR “connected manufacturing”[tiab] OR “industry 4.0”[tiab] OR “smart environment”[tiab] OR “digital twin”[tiab] OR “internet of things”[tiab] OR “control system”[tiab] OR “dependable system”[tiab] OR “supervisory control”[tiab] OR “dual control”[tiab] OR “aerospace”[tiab] OR “aero*”[tiab] OR “aeronautics”[tiab] OR “transport*”[tiab] OR “automotive”[tiab] OR “autonomous vehicle”[tiab] OR “smart cockpit”[tiab] OR “air traffic control”[tiab] OR “autonomous operation”[tiab] OR “automated driving”[tiab] OR “smart autonomous vehicle system”[tiab] OR “automat*”[tiab] OR “digital automat*”[tiab] OR “digital assistance system”[tiab] OR “automated decision aid”[tiab] OR “cloud computing”[tiab] OR “cognitive computing”[tiab] OR “enterprise system”[tiab] OR “information system”[tiab] OR “cognitive systems engineering”[tiab] OR “agent based system”[tiab] OR “ambient intelligence”[tiab] OR “one-to-many system”[tiab] OR “human-systems”[tiab] OR “human-machine system”[tiab] OR “human-autonomy-teaming”[tiab] OR “cognitive assistant”[tiab] OR “intelligent assistant”[tiab] OR “virtual assistant”[tiab] OR “virtual agent”[tiab] OR “synthetic teammate”[tiab] OR “intelligent human-machine interaction”[tiab] OR “cognitive assistance system”[tiab] OR “emotional-based agent”[tiab] OR “robotics”[tiab] OR “cyber-physical system”[tiab] OR “industrial robot”[tiab] OR “social robot”[tiab] OR “evolutionary robotics”[tiab] OR “cognitive robotics”[tiab] OR “aerial robotic”[tiab] OR “teleoperation”[tiab] OR “telerobotics”[tiab] OR “cyber-physical-human-system”[tiab] OR “human-cyber-physical system”[tiab] OR “human robot interaction”[tiab] OR “human machine interaction”[tiab] OR “collaborative robot”[tiab] OR “cobot”[tiab] OR “physical human-robot interaction”[tiab] OR “physical-robot-human interaction”[tiab] OR “shared robotic task”[tiab] OR “human-robot team”[tiab] OR “closed-loop human-robot interaction”[tiab] OR “real-time human-robot interaction”[tiab] OR “robotic symbiotic network”[tiab] OR “human-robot collaboration”[tiab] OR “safe physical human–robot collaboration”[tiab] OR “brain mediated human-robot interaction”[tiab] OR “admittance control”[tiab] OR

		“distance education”[tiab] OR “cognitive medical robot”[tiab] OR “smart health environment”[tiab] OR “medical robots and system”[tiab] OR “robot assisted surgery”[tiab] OR “robotic surgical procedures”[tiab] OR “computer-aided diagnosis”[tiab] OR “neuronavigation”[tiab]) AND (“cognitive state”[tiab] OR “mental state”[tiab] OR “psychological state”[tiab] OR “mental process”[tiab] OR “mental condition”[tiab] OR “cognitive function”[tiab] OR “cognitive mental state”[tiab] OR “cognitive processes”[tiab] OR “psychological state”[tiab] OR “cognition”[tiab] OR “emotion”[tiab] OR “engagement”[tiab] OR “workload”[tiab] OR “mental workload”[tiab] OR “situational awareness”[tiab] OR “multitasking”[tiab] OR “attention”[tiab] OR “drowsiness”[tiab] OR “distraction”[tiab] OR “alertness”[tiab] OR “fatigue”[tiab] OR “boredom”[tiab] OR “anxiety”[tiab] OR “stress”[tiab] OR “emotion”[tiab] OR “vigilance”[tiab] OR “working memory”[tiab] OR “intent”[tiab] OR “distraction”[tiab] OR “alertness”[tiab] OR “confusion”[tiab] OR “human intention”[tiab] OR “cognitive absorption”[tiab] OR “information overload”[tiab] OR “cognitive readiness”[tiab] OR “sleepiness”[tiab] OR “attentional tunneling”[tiab] OR “vigilance”[tiab] OR “cognitive workload”[tiab] OR “inattention”[tiab]) AND (“adaptive”[tiab] OR “adapt*”[tiab] OR “adaptive automation”[tiab] OR “adaptive systems”[tiab] OR “assistive”[tiab] OR “inclusive design”[tiab] OR “human-automation performance”[tiab] OR “real-time adaptive system”[tiab] OR “adaptable automation”[tiab] OR “dynamic function allocation”[tiab] OR “fallback”[tiab] OR “adaptive assistance”[tiab] OR “dynamic adaptation”[tiab] OR “adaptive human-automation systems”[tiab] OR “human in-the-loop”[tiab] OR “adaptive interface”[tiab] OR “cognitive automation”[tiab] OR “adaptive learning”[tiab] OR “adaptive control”[tiab] OR “system adaptation”[tiab] OR “adaptive workload allocation”[tiab] OR “adaptive aiding”[tiab] OR “adaptive cruise control”[tiab] OR “flexible automation”[tiab] OR “adaptive mitigation strategies”[tiab] OR “human-systems inclusion”[tiab] OR “human-systems integration”[tiab] OR “multimodal interaction”[tiab] OR “biofeedback”[tiab] OR “augmented cognition”[tiab] OR “psychological adjusting”[tiab] OR “biofeedback”[tiab] OR “cognitive performance enhancement”[tiab] OR “cognitive enhancement”[tiab] OR “mental support”[tiab] OR “neurofeedback”[tiab] OR “cognitive monitoring”[tiab]) NOT (“Plant”[tiab] OR “Rehabilitation”[tiab] OR “Disease”[tiab] OR “Therapy”[tiab] OR “Therapeutic”[tiab] OR “Microbiology”[tiab] OR “Microbial”[tiab] OR “Pathogen”[tiab] OR “Aerobic”[tiab] OR “aerodynamic”[tiab] OR “aerosol”[tiab] OR “aerogenes”[tiab] OR “molecular”[tiab] OR “protein*” [tiab]) AND (2012:2022[pdat]))
ACM	3155	[[Abstract: "manifactur*"] OR [Abstract: "smart manifactur*"] OR [Abstract: "smart factory"] OR [Abstract: "connected manufacturing"] OR [Abstract: "industry 4.0"] OR [Abstract: "smart environment"] OR [Abstract: "digital twin"] OR [Abstract: "internet of things"] OR [Abstract: "control system"] OR [Abstract: "dependable system"] OR [Abstract: "supervisory control"] OR [Abstract: "dual control"] OR [Abstract: "aerospace"] OR [Abstract: "aero*"] OR [Abstract: "aeronautics"] OR [Abstract: "transport*"] OR [Abstract: "automotive"] OR [Abstract: "autonomous vehicle"] OR [Abstract: "smart cockpit"] OR [Abstract: "air traffic control"] OR [Abstract: "autonomous operation"] OR [Abstract: "automated driving"] OR [Abstract: "smart autonomous vehicle system"] OR [Abstract: "automat*"] OR [Abstract: "digital automat*"] OR

		[Abstract: "digital assistance system"] OR [Abstract: "automated decision aid"] OR [Abstract: " cloud computing"] OR [Abstract: "cognitive computing"] OR [Abstract: "enterprise system"] OR [Abstract: "information system"] OR [Abstract: "cognitive systems engineering"] OR [Abstract: "agent based system"] OR [Abstract: "ambient intelligence"] OR [Abstract: "one-to-many system"] OR [Abstract: "human-systems"] OR [Abstract: "human-machine system"] OR [Abstract: "human-autonomy-teaming"] OR [Abstract: "cognitive assistant"] OR [Abstract: "intelligent assistant"] OR [Abstract: "virtual assistant"] OR [Abstract: "virtual agent"] OR [Abstract: "synthetic teammate"] OR [Abstract: "intelligent human-machine interaction"] OR [Abstract: "cognitive assistance system"] OR [Abstract: "emotional-based agent"] OR [Abstract: "robotics"] OR [Abstract: "cyber-physical system"] OR [Abstract: "industrial robot"] OR [Abstract: "social robot"] OR [Abstract: "evolutionary robotics"] OR [Abstract: "cognitive robotics"] OR [Abstract: "aerial robotic"] OR [Abstract: "teleoperation"] OR [Abstract: "telerobotics"] OR [Abstract: "cyber-physical-human-system"] OR [Abstract: "human-cyber-physical system"] OR [Abstract: "human robot interaction"] OR [Abstract: "human machine interaction"] OR [Abstract: "collaborative robot"] OR [Abstract: "cobot"] OR [Abstract: "physical human-robot interaction"] OR [Abstract: "physical-robot-human interaction"] OR [Abstract: "shared robotic task"] OR [Abstract: "human-robot team"] OR [Abstract: "closed-loop human-robot interaction"] OR [Abstract: "real-time human-robot interaction"] OR [Abstract: "robotic symbiotic network"] OR [Abstract: "human-robot collaboration"] OR [Abstract: "safe physical human–robot collaboration"] OR [Abstract: "brain mediated human-robot interaction"] OR [Abstract: "admittance control"] OR [Abstract: "distance education"] OR [Abstract: "cognitive medical robot"] OR [Abstract: "smart health environment"] OR [Abstract: "medical robots and system"] OR [Abstract: "robot assisted surgery"] OR [Abstract: "robotic surgical procedures"] OR [Abstract: "computer-aided diagnosis"] OR [Abstract: "neuronavigation"]] AND [[Full Text: "cognitive state"] OR [Full Text: "mental state"] OR [Full Text: "psychological state"] OR [Full Text: "mental process"] OR [Full Text: "mental condition"] OR [Full Text: "cognitive function"] OR [Full Text: "cognitive mental state"] OR [Full Text: "cognitive processes"] OR [Full Text: "psychological state"] OR [Full Text: "cognition"] OR [Full Text: "emotion"] OR [Full Text: "engagement"] OR [Full Text: "workload"] OR [Full Text: "mental workload"] OR [Full Text: "situational awareness"] OR [Full Text: "multitasking"] OR [Full Text: "attention"] OR [Full Text: "drowsiness"] OR [Full Text: "distraction"] OR [Full Text: "alertness"] OR [Full Text: "fatigue"] OR [Full Text: "boredom"] OR [Full Text: "anxiety"] OR [Full Text: "stress"] OR [Full Text: "emotion"] OR [Full Text: "vigilance"] OR [Full Text: "working memory"] OR [Full Text: "intent"] OR [Full Text: "distraction"] OR [Full Text: "alertness"] OR [Full Text: "confusion"] OR [Full Text: "human intention"] OR [Full Text: "cognitive absorption"] OR [Full Text: "information overload"] OR [Full Text: "cognitive readiness"] OR [Full Text: "sleepiness"] OR [Full Text: "attentional tunneling"] OR [Full Text: "vigilance"] OR [Full Text: "cognitive workload"] OR [Full Text: "inattention"]] AND [[Full Text: "adaptive"] OR [Full Text: "adapt*"] OR [Full Text: "adaptive automation"] OR [Full Text: "adaptive systems"] OR [Full Text: "assistive"] OR [Full Text: "inclusive design"] OR [Full Text: "human-automation performance"] OR [Full Text: "real-time adaptive system"] OR [Full Text: "adaptable automation"] OR [Full Text: "dynamic function allocation"] OR [Full Text: "fallback"] OR [Full Text:

		"adaptive assistance"] OR [Full Text: "dynamic adaptation"] OR [Full Text: "adaptive human-automation systems"] OR [Full Text: "human in-the-loop"] OR [Full Text: "adaptive interface"] OR [Full Text: "cognitive automation"] OR [Full Text: "adaptive learning"] OR [Full Text: "adaptive control"] OR [Full Text: "system adaptation"] OR [Full Text: "adaptive workload allocation"] OR [Full Text: "adaptive aiding"] OR [Full Text: "adaptive cruise control"] OR [Full Text: "flexible automation"] OR [Full Text: "adaptive mitigation strategies"] OR [Full Text: "human-systems inclusion"] OR [Full Text: "human-systems integration"] OR [Full Text: "multimodal interaction"] OR [Full Text: "biofeedback"] OR [Full Text: "augmented cognition"] OR [Full Text: "psychological adjusting"] OR [Full Text: "biofeedback"] OR [Full Text: "cognitive performance enhancement"] OR [Full Text: "cognitive enhancement"] OR [Full Text: "mental support"] OR [Full Text: "neurofeedback"] OR [Full Text: "cognitive monitoring"]] AND NOT [[Abstract: "plant"] OR [Abstract: "rehabilitation"] OR [Abstract: "disease"] OR [Abstract: "therapy"] OR [Abstract: "therapeutic"] OR [Abstract: "microbiology"] OR [Abstract: "microbial"] OR [Abstract: "pathogen"] OR [Abstract: "aerobic"] OR [Abstract: "aerodynamic"] OR [Abstract: "aerosol"] OR [Abstract: "aerogenes"]] AND [Publication Date: (01/01/2012 TO 12/31/2022)]

A1.2.2 Final queries

Data was extracted from the different database using as basis the Web of Science query. Query was adapted based on the database, Abstract/Title/Key word for the first part of the keyword, for the AI related keyword full text if available. 2012 to 2022 filter was set and when option for peer review filtering was available the option was used.

Table 39

Search Phase – Final queries

WoS	2242	TS=(“manifactur*” OR “smart manifactur*” OR “smart factory” OR “connected manufacturing” OR “industry 4.0” OR “smart environment” OR “digital twin” OR “internet of things” OR “control system” OR “dependable system” OR “supervisory control” OR “dual control” OR “aerospace” OR “aero*” OR “aeronautics” OR “transport*” OR “automotive” OR “autonomous vehicle” OR “smart cockpit” OR “air traffic control” OR “autonomous operation” OR “automated driving” OR “smart autonomous vehicle system” OR “automat*” OR “digital automat*” OR “digital assistance system” OR “automated decision aid” OR “ cloud computing” OR “cognitive computing” OR “enterprise system” OR “information system” OR “cognitive systems engineering” OR “agent based system” OR “ambient intelligence” OR “one-to-many system” OR “human-systems” OR “human-machine system” OR “human-autonomy-teaming” OR “cognitive assistant” OR “intelligent assistant” OR “virtual assistant” OR “virtual agent” OR “synthetic teammate” OR “intelligent human-machine interaction” OR “cognitive assistance system” OR “emotional-based agent” OR “robotics” OR “cyber-physical system” OR “industrial robot” OR “social robot” OR “evolutionary robotics” OR “cognitive robotics” OR “aerial robotic” OR “teleoperation” OR “telerobotics” OR “cyber-physical-human-system” OR “human-cyber-physical system” OR “human robot interaction” OR “human machine interaction” OR “collaborative robot” OR “cobot” OR “physical human-robot interaction” OR “physical-robot-human interaction” OR “shared robotic task” OR “human-robot team” OR “closed-loop human-robot interaction” OR “real-time human-robot interaction” OR “robotic symbiotic network” OR “human-robot collaboration” OR “safe physical human–robot collaboration” OR “brain mediated human-robot interaction” OR “admittance control” OR “distance education” OR “cognitive medical robot” OR “smart health environment” OR “medical robots and system” OR “robot assisted surgery” OR “robotic surgical procedures” OR “computer-aided diagnosis” OR “neuronavigation”) AND (TS=(“cognitive state” OR “mental state” OR “psychological state” OR “mental process” OR “mental condition” OR “cognitive function” OR “cognitive mental state” OR “cognitive processes” OR “psychological state” OR “cognition” OR “emotion” OR “engagement” OR “workload” OR “mental workload” OR “situational awareness” OR “multitasking” OR “attention” OR “drowsiness” OR

		“distraction” OR “alertness” OR “fatigue” OR “boredom” OR “anxiety” OR “stress” OR “emotion” OR “vigilance” OR “working memory” OR “intent” OR “distraction” OR “alertness” OR “confusion” OR “human intention” OR “cognitive absorption” OR “information overload” OR “cognitive readiness” OR “sleepiness” OR “attentional tunneling” OR “vigilance” OR “cognitive workload” OR “inattention”) AND TS=(“adaptive” OR “adapt*” OR “adaptive automation” OR “adaptive systems” OR “assistive” OR “inclusive design” OR “human-automation performance” OR “real-time adaptive system” OR “adaptable automation” OR “dynamic function allocation” OR “fallback” OR “adaptive assistance” OR “dynamic adaptation” OR “adaptive human-automation systems” OR “human in-the-loop” OR “adaptive interface” OR “cognitive automation” OR “adaptive learning” OR “adaptive control” OR “system adaptation” OR “adaptive workload allocation” OR “adaptive aiding” OR “adaptive cruise control” OR “flexible automation” OR “adaptive mitigation strategies” OR “human-systems inclusion” OR “human-systems integration” OR “multimodal interaction” OR “biofeedback” OR “augmented cognition” OR “psychological adjusting” OR “biofeedback” OR “cognitive performance enhancement” OR “cognitive enhancement” OR “mental support” OR “neurofeedback” OR “cognitive monitoring”)) AND TS=(“AI” OR “artificial intelligence” OR “digital intelligence” OR “machine learning” OR “analytics” OR “big data” OR “machine intelligence” OR “deep learning” OR “prediction model” OR “neural network” OR “computational intelligence” OR “predictive models” OR “affective computing” OR “physiological computing” OR “computer reasoning” OR “decision support systems” OR “big data” OR “expert systems” OR “statistical learning” OR “artificial consciousness” OR “machine consciousness” OR “wearable computing” OR “supervised learning” OR “unsupervised learning” OR “reinforcement learning” OR “artificial neural network” OR “artificial cognition” OR “machine perception” OR “attention-based computing” OR “computer vision” OR “support vector machines” OR “natural language processing” OR “convolutional neural network” OR “convolutional network” OR “computer heuristics” OR “regression” OR “transfer learning” OR “genetic algorithm” OR “recommender system” OR “recommendation algorithm” OR “random forest” OR “linear classifier” OR “nearest neighbor search” OR “decision tree” OR “hidden markov chain” OR “ensemble learning” OR “relevance vector machine” OR “brain-computer interface” OR “brain-machine interface” OR “neurotechnology”) NOT TS=(“Plant” OR “Rehabilitation” OR “Disease” OR “Therapy” OR “Therapeutic” OR “Microbiology” OR “Microbial” OR “Pathogen” OR “Aerobic” OR “aerodynamic” OR “aerosol” OR “aerogenes”)
Scopus	4859	((TITLE-ABS-KEY ("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "smart environment" OR "digital twin" OR "internet of things" OR "control system" OR "dependable system" OR "supervisory control" OR "dual control" OR "aerospace" OR "aero*" OR "aeronautics" OR "transport*" OR "automotive" OR "autonomous vehicle" OR "smart cockpit" OR "air traffic control" OR "autonomous operation" OR "automated driving" OR "smart autonomous vehicle system" OR "automat*" OR "digital automat*" OR "digital assistance system" OR "automated decision aid" OR " cloud computing" OR "cognitive computing" OR "enterprise system" OR "information system" OR "cognitive systems engineering" OR "agent based system" OR "ambient intelligence" OR "one-to-many system" OR "human-

		systems" OR "human-machine system" OR "human-autonomy-teaming" OR "cognitive assistant" OR "intelligent assistant" OR "virtual assistant" OR "virtual agent" OR "synthetic teammate" OR "intelligent human-machine interaction" OR "cognitive assistance system" OR "emotional-based agent" OR "robotics" OR "cyber-physical system" OR "industrial robot" OR "social robot" OR "evolutionary robotics" OR "cognitive robotics" OR "aerial robotic" OR "teleoperation" OR "telerobotics" OR "cyber-physical-human-system" OR "human-cyber-physical system" OR "human robot interaction" OR "human machine interaction" OR "collaborative robot" OR "cobot" OR "physical human-robot interaction" OR "physical-robot-human interaction" OR "shared robotic task" OR "human-robot team" OR "closed-loop human-robot interaction" OR "real-time human-robot interaction" OR "robotic symbiotic network" OR "human-robot collaboration" OR "safe physical human--robot collaboration" OR "brain mediated human-robot interaction" OR "admittance control" OR "distance education" OR "cognitive medical robot" OR "smart health environment" OR "medical robots and system" OR "robot assisted surgery" OR "robotic surgical procedures" OR "computer-aided diagnosis" OR "neuronavigation")) AND (TITLE-ABS-KEY ("cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition" OR "cognitive function" OR "cognitive mental state" OR "cognitive processes" OR "psychological state" OR "cognition" OR "emotion" OR "engagement" OR "workload" OR "mental workload" OR "situational awareness" OR "multitasking" OR "attention" OR "drowsiness" OR "distraction" OR "alertness" OR "fatigue" OR "boredom" OR "anxiety" OR "stress" OR "emotion" OR "vigilance" OR "working memory" OR "intent" OR "distraction" OR "alertness" OR "confusion" OR "human intention" OR "cognitive absorption" OR "information overload" OR "cognitive readiness" OR "sleepiness" OR "attentional tunneling" OR "vigilance" OR "cognitive workload" OR "inattention")) AND (TITLE-ABS-KEY ("adaptive" OR "adapt*" OR "adaptive automation" OR "adaptive systems" OR "assistive" OR "inclusive design" OR "human-automation performance" OR "real-time adaptive system" OR "adaptable automation" OR "dynamic function allocation" OR "fallback" OR "adaptive assistance" OR "dynamic adaptation" OR "adaptive human-automation systems" OR "human in-the-loop" OR "adaptive interface" OR "cognitive automation" OR "adaptive learning" OR "adaptive control" OR "system adaptation" OR "adaptive workload allocation" OR "adaptive aiding" OR "adaptive cruise control" OR "flexible automation" OR "adaptive mitigation strategies" OR "human-systems inclusion" OR "human-systems integration" OR "multimodal interaction" OR "biofeedback" OR "augmented cognition" OR "psychological adjusting" OR "biofeedback" OR "cognitive performance enhancement" OR "cognitive enhancement" OR "mental support" OR "neurofeedback" OR "cognitive monitoring")) AND (TITLE-ABS-KEY (“AI” OR “artificial intelligence” OR “digital intelligence” OR “machine learning” OR “analytics” OR “big data” OR “machine intelligence” OR “deep learning” OR “prediction model” OR “neural network” OR “computational intelligence” OR “predictive models” OR “affective computing” OR “physiological computing” OR “computer reasoning” OR “decision support systems” OR “big data” OR “expert systems” OR “statistical learning” OR “artificial consciousness” OR “machine consciousness” OR “wearable computing” OR “supervised learning” OR “unsupervised learning” OR “reinforcement learning” OR “artificial neural network” OR “artificial cognition” OR “machine perception” OR “attention-based computing”

		OR “computer vision” OR “support vector machines” OR “natural language processing” OR “convolutional neural network” OR “convolutional network” OR “computer heuristics” OR “regression” OR “transfer learning” OR “genetic algorithm” OR “recommender system” OR “recommendation algorithm” OR “random forest” OR “linear classifier” OR “nearest neighbor search” OR “decision tree” OR “hidden markov chain” OR “ensemble learning” OR “relevance vector machine” OR “brain-computer interface” OR “brain-machine interface” OR “neurotechnology”)) AND NOT (TITLE-ABS-KEY ("Plant" OR "Rehabilitation" OR "Disease" OR "Therapy" OR "Therapeutic" OR "Microbiology" OR "Microbial" OR "Pathogen" OR "Aerobic" OR "aerodynamic" OR "aerosol" OR "aerogenes"))) AND PUBYEAR > 2011 AND (EXCLUDE (SUBJAREA,"BIOC"))
PubMed	395	((“manifactur*”[tiab] OR “smart manifactur*”[tiab] OR “smart factory”[tiab] OR “connected manufacturing”[tiab] OR “industry 4.0”[tiab] OR “smart environment”[tiab] OR “digital twin”[tiab] OR “internet of things”[tiab] OR “control system”[tiab] OR “dependable system”[tiab] OR “supervisory control”[tiab] OR “dual control”[tiab] OR “aerospace”[tiab] OR “aero*”[tiab] OR “aeronautics”[tiab] OR “transport*”[tiab] OR “automotive”[tiab] OR “autonomous vehicle”[tiab] OR “smart cockpit”[tiab] OR “air traffic control”[tiab] OR “autonomous operation”[tiab] OR “automated driving”[tiab] OR “smart autonomous vehicle system”[tiab] OR “automat*”[tiab] OR “digital automat*”[tiab] OR “digital assistance system”[tiab] OR “automated decision aid”[tiab] OR “cloud computing”[tiab] OR “cognitive computing”[tiab] OR “enterprise system”[tiab] OR “information system”[tiab] OR “cognitive systems engineering”[tiab] OR “agent based system”[tiab] OR “ambient intelligence”[tiab] OR “one-to-many system”[tiab] OR “human-systems”[tiab] OR “human-machine system”[tiab] OR “human-autonomy-teaming”[tiab] OR “cognitive assistant”[tiab] OR “intelligent assistant”[tiab] OR “virtual assistant”[tiab] OR “virtual agent”[tiab] OR “synthetic teammate”[tiab] OR “intelligent human-machine interaction”[tiab] OR “cognitive assistance system”[tiab] OR “emotional-based agent”[tiab] OR “robotics”[tiab] OR “cyber-physical system”[tiab] OR “industrial robot”[tiab] OR “social robot”[tiab] OR “evolutionary robotics”[tiab] OR “cognitive robotics”[tiab] OR “aerial robotic”[tiab] OR “teleoperation”[tiab] OR “telerobotics”[tiab] OR “cyber-physical-human-system”[tiab] OR “human-cyber-physical system”[tiab] OR “human robot interaction”[tiab] OR “human machine interaction”[tiab] OR “collaborative robot”[tiab] OR “cobot”[tiab] OR “physical human-robot interaction”[tiab] OR “physical-robot-human interaction”[tiab] OR “shared robotic task”[tiab] OR “human-robot team”[tiab] OR “closed-loop human-robot interaction”[tiab] OR “real-time human-robot interaction”[tiab] OR “robotic symbiotic network”[tiab] OR “human-robot collaboration”[tiab] OR “safe physical human–robot collaboration”[tiab] OR “brain mediated human-robot interaction”[tiab] OR “admittance control”[tiab] OR “distance education”[tiab] OR “cognitive medical robot”[tiab] OR “smart health environment”[tiab] OR “medical robots and system”[tiab] OR “robot assisted surgery”[tiab] OR “robotic surgical procedures”[tiab] OR “computer-aided diagnosis”[tiab] OR “neuronavigation”[tiab]) AND (“cognitive state”[tiab] OR “mental state”[tiab] OR “psychological state”[tiab] OR “mental process”[tiab] OR “mental condition”[tiab] OR “cognitive function”[tiab] OR “cognitive mental state”[tiab] OR “cognitive

		processes"[tiab] OR "psychological state"[tiab] OR "cognition"[tiab] OR "emotion"[tiab] OR "engagement"[tiab] OR "workload"[tiab] OR "mental workload"[tiab] OR "situational awareness"[tiab] OR "multitasking"[tiab] OR "attention"[tiab] OR "drowsiness"[tiab] OR "distraction"[tiab] OR "alertness"[tiab] OR "fatigue"[tiab] OR "boredom"[tiab] OR "anxiety"[tiab] OR "stress"[tiab] OR "emotion"[tiab] OR "vigilance"[tiab] OR "working memory"[tiab] OR "intent"[tiab] OR "distraction"[tiab] OR "alertness"[tiab] OR "confusion"[tiab] OR "human intention"[tiab] OR "cognitive absorption"[tiab] OR "information overload"[tiab] OR "cognitive readiness"[tiab] OR "sleepiness"[tiab] OR "attentional tunneling"[tiab] OR "vigilance"[tiab] OR "cognitive workload"[tiab] OR "inattention"[tiab]) AND ("adaptive"[tiab] OR "adapt*"[tiab] OR "adaptive automation"[tiab] OR "adaptive systems"[tiab] OR "assistive"[tiab] OR "inclusive design"[tiab] OR "human-automation performance"[tiab] OR "real-time adaptive system"[tiab] OR "adaptable automation"[tiab] OR "dynamic function allocation"[tiab] OR "fallback"[tiab] OR "adaptive assistance"[tiab] OR "dynamic adaptation"[tiab] OR "adaptive human-automation systems"[tiab] OR "human in-the-loop"[tiab] OR "adaptive interface"[tiab] OR "cognitive automation"[tiab] OR "adaptive learning"[tiab] OR "adaptive control"[tiab] OR "system adaptation"[tiab] OR "adaptive workload allocation"[tiab] OR "adaptive aiding"[tiab] OR "adaptive cruise control"[tiab] OR "flexible automation"[tiab] OR "adaptive mitigation strategies"[tiab] OR "human-systems inclusion"[tiab] OR "human-systems integration"[tiab] OR "multimodal interaction"[tiab] OR "biofeedback"[tiab] OR "augmented cognition"[tiab] OR "psychological adjusting"[tiab] OR "biofeedback"[tiab] OR "cognitive performance enhancement"[tiab] OR "cognitive enhancement"[tiab] OR "mental support"[tiab] OR "neurofeedback"[tiab] OR "cognitive monitoring"[tiab]) AND ("AI"[All Fields] OR "artificial intelligence"[All Fields] OR "digital intelligence"[All Fields] OR "machine learning"[All Fields] OR "analytics"[All Fields] OR "big data"[All Fields] OR "machine intelligence"[All Fields] OR "deep learning"[All Fields] OR "prediction model"[All Fields] OR "neural network"[All Fields] OR "computational intelligence"[All Fields] OR "predictive models"[All Fields] OR "affective computing"[All Fields] OR "physiological computing"[All Fields] OR "computer reasoning"[All Fields] OR "decision support systems"[All Fields] OR "big data"[All Fields] OR "expert systems"[All Fields] OR "statistical learning"[All Fields] OR "artificial consciousness"[All Fields] OR "machine consciousness"[All Fields] OR "wearable computing"[All Fields] OR "supervised learning"[All Fields] OR "unsupervised learning"[All Fields] OR "reinforcement learning"[All Fields] OR "artificial neural network"[All Fields] OR "artificial cognition"[All Fields] OR "machine perception"[All Fields] OR "attention-based computing"[All Fields] OR "computer vision"[All Fields] OR "support vector machines"[All Fields] OR "natural language processing"[All Fields] OR "convolutional neural network"[All Fields] OR "convolutional network"[All Fields] OR "computer heuristics"[All Fields] OR "regression"[All Fields] OR "transfer learning"[All Fields] OR "genetic algorithm"[All Fields] OR "recommender system"[All Fields] OR "recommendation algorithm"[All Fields] OR "random forest"[All Fields] OR "linear classifier"[All Fields] OR "nearest neighbor search"[All Fields] OR "decision tree"[All Fields] OR "hidden markov chain"[All Fields] OR "ensemble learning"[All Fields] OR "relevance vector machine"[All Fields] OR "brain-computer interface"[All Fields] OR "brain-machine interface"[All Fields]

		OR “neurotechnology”[All Fields]) NOT (“Plant”[tiab] OR “Rehabilitation”[tiab] OR “Disease”[tiab] OR “Therapy”[tiab] OR “Therapeutic”[tiab] OR “Microbiology”[tiab] OR “Microbial”[tiab] OR “Pathogen”[tiab] OR “Aerobic”[tiab] OR “aerodynamic”[tiab] OR “aerosol”[tiab] OR “aerogenes”[tiab] OR “molecular”[tiab] OR “protein*” [tiab]) AND (2012:2022[pdat]))
Proquest	185	ab(("manifactur*" OR "smart manifactur*" OR "smart factory" OR "connected manufacturing" OR "industry 4.0" OR "smart environment" OR "digital twin" OR "internet of things" OR "control system" OR "dependable system" OR "supervisory control" OR "dual control" OR "aerospace" OR "aero*" OR "aeronautics" OR "transport*" OR "automotive" OR "autonomous vehicle" OR "smart cockpit" OR "air traffic control" OR "autonomous operation" OR "automated driving" OR "smart autonomous vehicle system" OR "automat*" OR "digital automat*" OR "digital assistance system" OR "automated decision aid" OR " cloud computing" OR "cognitive computing" OR "enterprise system" OR "information system" OR "cognitive systems engineering" OR "agent based system" OR "ambient intelligence" OR "one-to-many system" OR "human-systems" OR "human-machine system" OR "human-autonomy-teaming" OR "cognitive assistant" OR "intelligent assistant" OR "virtual assistant" OR "virtual agent" OR "synthetic teammate" OR "intelligent human-machine interaction" OR "cognitive assistance system" OR "emotional-based agent" OR "robotics" OR "cyber-physical system" OR "industrial robot" OR "social robot" OR "evolutionary robotics" OR "cognitive robotics" OR "aerial robotic" OR "teleoperation" OR "telerobotics" OR "cyber-physical-human-system" OR "human-cyber-physical system" OR "human robot interaction" OR "human machine interaction" OR "collaborative robot" OR "cobot" OR "physical human-robot interaction" OR "physical-robot-human interaction" OR "shared robotic task" OR "human-robot team" OR "closed-loop human-robot interaction" OR "real-time human-robot interaction" OR "robotic symbiotic network" OR "human-robot collaboration" OR "safe physical human robot collaboration" OR "brain mediated human-robot interaction" OR "admittance control" OR "distance education" OR "cognitive medical robot" OR "smart health environment" OR "medical robots and system" OR "robot assisted surgery" OR "robotic surgical procedures" OR "computer-aided diagnosis" OR "neuronavigation")) AND ab(("cognitive state" OR "mental state" OR "psychological state" OR "mental process" OR "mental condition" OR "cognitive function" OR "cognitive mental state" OR "cognitive processes" OR "psychological state" OR "cognition" OR "emotion" OR "engagement" OR "workload" OR "mental workload" OR "situational awareness" OR "multitasking" OR "attention" OR "drowsiness" OR "distraction" OR "alertness" OR "fatigue" OR "boredom" OR "anxiety" OR "stress" OR "emotion" OR "vigilance" OR "working memory" OR "intent" OR "distraction" OR "alertness" OR "confusion" OR "human intention" OR "cognitive absorption" OR "information overload" OR "cognitive readiness" OR "sleepiness" OR "attentional tunneling" OR "vigilance" OR "cognitive workload" OR "inattention")) AND ab(("adaptive" OR "adapt*" OR "adaptive automation" OR "adaptive systems" OR "assistive" OR "inclusive design" OR "human-automation performance" OR "real-time adaptive system" OR "adaptable automation" OR "dynamic function allocation" OR "fallback" OR "adaptive assistance" OR "dynamic adaptation" OR "adaptive human-automation systems" OR

		"human in-the-loop" OR "adaptive interface" OR "cognitive automation" OR "adaptive learning" OR "adaptive control" OR "system adaptation" OR "adaptive workload allocation" OR "adaptive aiding" OR "adaptive cruise control" OR "flexible automation" OR "adaptive mitigation strategies" OR "human-systems inclusion" OR "human-systems integration" OR "multimodal interaction" OR "biofeedback" OR "augmented cognition" OR "psychological adjusting" OR "biofeedback" OR "cognitive performance enhancement" OR "cognitive enhancement" OR "mental support" OR "neurofeedback" OR "cognitive monitoring")) AND ("AI" OR "artificial intelligence" OR "digital intelligence" OR "machine learning" OR "analytics" OR "big data" OR "machine intelligence" OR "deep learning" OR "prediction model" OR "neural network" OR "computational intelligence" OR "predictive models" OR "affective computing" OR "physiological computing" OR "computer reasoning" OR "decision support systems" OR "big data" OR "expert systems" OR "statistical learning" OR "artificial consciousness" OR "machine consciousness" OR "wearable computing" OR "supervised learning" OR "unsupervised learning" OR "reinforcement learning" OR "artificial neural network" OR "artificial cognition" OR "machine perception" OR "attention-based computing" OR "computer vision" OR "support vector machines" OR "natural language processing" OR "convolutional neural network" OR "convolutional network" OR "computer heuristics" OR "regression" OR "transfer learning" OR "genetic algorithm" OR "recommender system" OR "recommendation algorithm" OR "random forest" OR "linear classifier" OR "nearest neighbor search" OR "decision tree" OR "hidden Markov chain" OR "ensemble learning" OR "relevance vector machine" OR "brain-computer interface" OR "brain-machine interface" OR "neurotechnology")
PsyInfo	150	(“manifactur*” OR “smart manifactur*” OR “smart factory” OR “connected manufacturing” OR “industry 4.0” OR “smart environment” OR “digital twin” OR “internet of things” OR “control system” OR “dependable system” OR “supervisory control” OR “dual control” OR “aerospace” OR “aero*” OR “aeronautics” OR “transport*” OR “automotive” OR “autonomous vehicle” OR “smart cockpit” OR “air traffic control” OR “autonomous operation” OR “automated driving” OR “smart autonomous vehicle system” OR “automat*” OR “digital automat*” OR “digital assistance system” OR “automated decision aid” OR “ cloud computing” OR “cognitive computing” OR “enterprise system” OR “information system” OR “cognitive systems engineering” OR “agent based system” OR “ambient intelligence” OR “one-to-many system” OR “human-systems” OR “human-machine system” OR “human-autonomy-teaming” OR “cognitive assistant” OR “intelligent assistant” OR “virtual assistant” OR “virtual agent” OR “synthetic teammate” OR “intelligent human-machine interaction” OR “cognitive assistance system” OR “emotional-based agent” OR “robotics” OR “cyber-physical system” OR “industrial robot” OR “social robot” OR “evolutionary robotics” OR “cognitive robotics” OR “aerial robotic” OR “teleoperation” OR “telerobotics” OR “cyber-physical-human-system” OR “human-cyber-physical system” OR “human robot interaction” OR “human machine interaction” OR “collaborative robot” OR “cobot” OR “physical human-robot interaction” OR “physical-robot-human interaction” OR “shared robotic task” OR “human-robot team” OR “closed-loop human-robot interaction” OR “real-time human-robot interaction” OR “robotic symbiotic network” OR “human-robot collaboration” OR “safe physical

		human–robot collaboration” OR “brain mediated human-robot interaction” OR “admittance control” OR “distance education” OR “cognitive medical robot” OR “smart health environment” OR “medical robots and system” OR “robot assisted surgery” OR “robotic surgical procedures” OR “computer-aided diagnosis” OR “neuronavigation”) AND Abstract: (“cognitive state” OR “mental state” OR “psychological state” OR “mental process” OR “mental condition” OR “cognitive function” OR “cognitive mental state” OR “cognitive processes” OR “psychological state” OR “cognition” OR “emotion” OR “engagement” OR “workload” OR “mental workload” OR “situational awareness” OR “multitasking” OR “attention” OR “drowsiness” OR “distraction” OR “alertness” OR “fatigue” OR “boredom” OR “anxiety” OR “stress” OR “emotion” OR “vigilance” OR “working memory” OR “intent” OR “distraction” OR “alertness” OR “confusion” OR “human intention” OR “cognitive absorption” OR “information overload” OR “cognitive readiness” OR “sleepiness” OR “attentional tunneling” OR “vigilance” OR “cognitive workload” OR “inattention”) AND Abstract: (“adaptive” OR “adapt*” OR “adaptive automation” OR “adaptive systems” OR “assistive” OR “inclusive design” OR “human-automation performance” OR “real-time adaptive system” OR “adaptable automation” OR “dynamic function allocation” OR “fallback” OR “adaptive assistance” OR “dynamic adaptation” OR “adaptive human-automation systems” OR “human in-the-loop” OR “adaptive interface” OR “cognitive automation” OR “adaptive learning” OR “adaptive control” OR “system adaptation” OR “adaptive workload allocation” OR “adaptive aiding” OR “adaptive cruise control” OR “flexible automation” OR “adaptive mitigation strategies” OR “human-systems inclusion” OR “human-systems integration” OR “multimodal interaction” OR “biofeedback” OR “augmented cognition” OR “psychological adjusting” OR “biofeedback” OR “cognitive performance enhancement” OR “cognitive enhancement” OR “mental support” OR “neurofeedback” OR “cognitive monitoring”) AND Any Field: (“AI” OR “artificial intelligence” OR “digital intelligence” OR “machine learning” OR “analytics” OR “big data” OR “machine intelligence” OR “deep learning” OR “prediction model” OR “neural network” OR “computational intelligence” OR “predictive models” OR “affective computing” OR “physiological computing” OR “computer reasoning” OR “decision support systems” OR “big data” OR “expert systems” OR “statistical learning” OR “artificial consciousness” OR “machine consciousness” OR “wearable computing” OR “supervised learning” OR “unsupervised learning” OR “reinforcement learning” OR “artificial neural network” OR “artificial cognition” OR “machine perception” OR “attention-based computing” OR “computer vision” OR “support vector machines” OR “natural language processing” OR “convolutional neural network” OR “convolutional network” OR “computer heuristics” OR “regression” OR “transfer learning” OR “genetic algorithm” OR “recommender system” OR “recommendation algorithm” OR “random forest” OR “linear classifier” OR “nearest neighbor search” OR “decision tree” OR “hidden markov chain” OR “ensemble learning” OR “relevance vector machine” OR “brain-computer interface” OR “brain-machine interface” OR “neurotechnology”) AND Peer-Reviewed Journals only AND Year: 2012 To 2022
IEEE	2147	("All Metadata":“manifactur*” OR "All Metadata":“smart manifactur*” OR "All Metadata":“industry 4.0” OR "All Metadata":“internet of things” OR "All Metadata":“aero*” OR "All Metadata":“transport*” OR "All Metadata":“autonomous

Explore		vehicle” OR "All Metadata":“smart cockpit” OR "All Metadata":“air traffic control” OR "All Metadata":“autonomous operation” OR "All Metadata":“automat*” OR "All Metadata":“information system” OR "All Metadata":“cognitive systems engineering” OR "All Metadata":“human-machine system” OR "All Metadata":“intelligent assistant” OR "All Metadata":“virtual assistant” OR "All Metadata":“virtual agent” OR "All Metadata":“robotics” OR "All Metadata":“cyber-physical system” OR "All Metadata":“teleoperation” OR "All Metadata":“human machine interaction” OR "All Metadata":“human-robot collaboration” OR "All Metadata":“education” OR "All Metadata":“medical robot” OR "All Metadata":“smart health environment” OR "All Metadata":“robot assisted surgery”) AND ("All Metadata":“cognitive state” OR "All Metadata":“mental state” OR "All Metadata":“psychological state” OR "All Metadata":“mental process” OR "All Metadata":“mental condition” OR "All Metadata":“cognitive function” OR "All Metadata":“cognitive mental state” OR "All Metadata":“cognitive processes” OR "All Metadata":“psychological state”) AND ("All Metadata":“adaptive” OR "All Metadata":“adapt*” OR "All Metadata":“adaptive automation” OR "All Metadata":“adaptive systems” OR "All Metadata":“assistive” OR "All Metadata":“inclusive design” OR "All Metadata":“human-automation performance” OR "All Metadata":“real-time adaptive system” OR "All Metadata":“adaptable automation” OR "All Metadata":“fallback” OR "All Metadata":“adaptive assistance” OR "All Metadata":“dynamic adaptation” OR "All Metadata":“adaptive human-automation systems” OR "All Metadata":“human in-the-loop” OR "All Metadata":“adaptive interface” OR "All Metadata":“biofeedback”) AND ("Full Text & Metadata":“AI” OR "Full Text & Metadata":“artificial intelligence” OR "Full Text & Metadata":“digital intelligence” OR "Full Text & Metadata":“machine learning” OR "Full Text & Metadata":“analytics” OR "Full Text & Metadata":“big data” OR "Full Text & Metadata":“machine intelligence” OR "Full Text & Metadata":“deep learning” OR "Full Text & Metadata":“prediction model” OR "Full Text & Metadata":“neural network” OR "Full Text & Metadata":“computational intelligence” OR "Full Text & Metadata":“predictive models” OR "Full Text & Metadata":“affective computing” OR "Full Text & Metadata":“physiological computing” OR "Full Text & Metadata":“computer reasoning” OR "Full Text & Metadata":“decision support systems” OR "Full Text & Metadata":“big data” OR "Full Text & Metadata":“expert systems” OR "Full Text & Metadata":“statistical learning” OR "Full Text & Metadata":“brain-computer interface” OR "Full Text & Metadata":“brain-machine interface” OR "Full Text & Metadata":“neurotechnology”)

A.1.3 Keywords

Table 40

Phase 1 – Keywords

Context of use	States	Objective	A.I.
manifactur*, smart manifactur*, smart factory, connected manufacturing, industry 4.0, aerospace, aero*, aeronautics, transport*, automotive, socio-technic*	Cognitive state, mental state, psychological state, mental process, mental condition	adaptive, adapt*, adaptive automation, assistive, inclusive design, support, human-systems, human-systems inclusion, human-systems integration, human-machine systems, control systems, monitoring,	Machine intelligence, deep learning, prediction model, neural network, support vector machines. natural language processing, computer vision, supervised learning, unsupervised learning, reinforcement learning, statistical learning, computational intelligence, computer reasoning, computer heuristics, expert systems

Table 41

Phase 2 – Keywords

Context of use	States	Objective	A.I.
manifactur*, smart manifactur*, smart factory, connected manufacturing, industry 4.0, aerospace, aero*, aeronautics, transport*, automotive, socio-technic*, Automat*, digital automat*, cloud computing, cognitive computing, enterprise systems, information system,	Cognitive state, mental state, psychological state, mental process, mental condition, cognitive function, cognitive mental state, implicit cognitive processes, psychological state, cognition, emotion	adaptive, adapt*, adaptive automation, adaptive systems, assistive, inclusive design, support, human-systems, human-systems inclusion, human-systems integration, human-machine systems, control systems,	AI, Artificial Intelligence, Digital Intelligence, Machine learning, analytics, big data, data mining, Machine intelligence, deep learning, prediction model, neural network, support vector machines. natural language processing, computer vision, supervised learning, unsupervised learning, reinforcement learning, statistical learning, computational intelligence, computer reasoning, computer heuristics, expert systems, brain computer interface, decision support, decisions support system, Closed-loop systems, regression, random forest, predictive models, Artificial neural network, Linear Classifier, Virtual agents, artificial companions, cognitive agents
	engagement, workload, mental workload, affect, Situational awareness, Decision making,	Dependable systems, Dependability, rehabilitation, augmented cognition, real-time ,	

	Multitasking, attention, drowsiness, distraction, alertness, fatigue, Boredom, Anxiety, stress, risk, emotion, Vigilance, Working memory, intent, distraction, alertness, confusion	learner modeling, human-automation performance, cognitive state profile, mitigation strategies, biofeedback, Real-time adaptive system, adaptable automation, dynamic function allocation, fallback, Human robot interaction, Robot assisted, psychological adjusting, biofeedback, cognitive performance enhancement, Adaptive assistance, Human-autonomy-teaming, mental support, human machine interaction, dynamic adaptation, safety, Supervisory control, Dual control, adaptive human-automation systems	

Table 42

Phase 2 – Keywords

Context of use	States	Objective	A.I.
manifactur*, smart manifactur*, smart factory, connected manufacturing, industry 4.0, aerospace, aero*, aeronautics, transport*, automotive, socio-technic*, Automat*, digital automat*, cloud computing, cognitive computing, enterprise systems, information system, robotics, distance education, medical, medication service, medical, autonomous vehicules, cyber-physical systems, cognitive systems engineering, digital assistance systems, adaptive instructional systems, agent based systems, Cognitive Medical Robots, Smart Environment, personalized medicine, Ambient Intelligence, smart health environment, industrial robots, autonomous operations, medical robots and systems, smart cockpit, digital twin, one-to-many systems	Cognitive state, mental state, psychological state, mental process, mental condition, cognitive function, cognitive mental state, implicit cognitive processes, psychological state, cognition, emotion	adaptive, adapt*, adaptive automation, adaptive systems, assistive, inclusive design, support, human-systems, human-systems inclusion, human-systems integration, human-machine systems, control systems, Dependable systems, Dependability, augmented cognition, real-time , learner modeling, human-automation performance, cognitive state profile, mitigation strategies, biofeedback, Real-time adaptive system, adaptable automation, dynamic function allocation, fallback, Human robot interaction, Robot assisted, psychological adjusting, biofeedback, cognitive performance enhancement, Adaptive assistance, Human-autonomy-teaming, mental support, human machine interaction, dynamic adaptation, safety, Supervisory control, Dual control, adaptive human-automation systems, grasp planning, collaborative robots,	AI, Artificial Intelligence, Digital Intelligence, Machine learning, analytics, big data, data mining, Machine intelligence, deep learning, prediction model, neural network, support vector machines. natural language processing, computer vision, supervised learning, unsupervised learning, reinforcement learning, statistical learning, computational intelligence, computer reasoning, computer heuristics, expert systems, brain computer interface, decision support, decisions support system, Closed-loop systems, regression, random forest, predictive models, Artificial neural network, Linear Classifier, Virtual agents, artificial companions, cognitive agents, neurotechnologies, shoelace pattern, nearest neighbor search, optimized extreme learning machine, robust variabtional mode decomposition, Whale Optimization Algorithm, convultion network, human-

		cobots, admittance control, computer-aided diagnosis, social robots, Robot Assisted Training, Assistive Robotics, supervisory control, human in-the-loop, Cognitive assistant, intelligent assistant, virtual assistant, virtual agent, adaptive interface, neurofeedback, driver analyzer, driver model, cognitive automation, intuitive cognition, embodied cognition, adaptive learning, cognitive monitoring	autonomy teaming, synthethic teammate, affective computing, aspect-oriented convolutional neural network , artificial cognition, Autonomous Adaptive Intelligence, Transfer learning; genetic algorithm, Recommender system, fuzzy system, decision tree
vehicles, plane, jet, train, car, machine	engagement, workload, mental workload, affect, Situational awareness, Decision making, Multitasking, attention, drowsiness, distraction, alertness, fatigue, Boredom, Anxiety, stress, risk, emotion, Vigilance, Working memory, intent, distraction, alertness, confusion, human intention, cognitive absorption, mental diseases, information overload, cognitive readiness		

Table 43

Phase 4 – Keywords

Context of use	States	Objective	A.I.
manifactur*, smart manifactur*, smart factory, connected manufacturing, industry 4.0, smart environment, digital twin, internet of things, control system, dependable system, supervisory control, dual control, aerospace, aero*, aeronautics, transport*, automotive, autonomous vehicle, smart cockpit, air traffic control, autonomous operation, automated driving, smart autonomous vehicle system, automat*, digital automat*, digital assistance system, automated decision aid, cloud computing, cognitive computing, enterprise system, information system, cognitive systems engineering, agent based system, ambient intelligence, one-to-many system, human-systems, human-machine system, human-autonomy-teaming, cognitive assistant, intelligent assistant, virtual assistant, virtual agent, synthetic teammate, intelligent human-machine interaction, cognitive	cognitive state, mental state, psychological state, mental process, mental condition, cognitive function, cognitive mental state, cognitive processes, psychological state, cognition, emotion, engagement, workload, mental workload, situational awareness, multitasking, attention, drowsiness, distraction, alertness, fatigue, boredom, anxiety, stress, emotion, vigilance, working memory, intent, distraction, alertness, confusion, human intention, cognitive absorption, information overload, cognitive readiness, sleepiness, attentional tunneling, vigilance, cognitive workload, inattention	adaptive, adapt*, adaptive automation, adaptive systems, assistive, inclusive design, human-automation performance, real-time adaptive system, adaptable automation, dynamic function allocation, fallback, adaptive assistance, dynamic adaptation, adaptive human-automation systems, human in-the-loop, adaptive interface, cognitive automation, adaptive learning, adaptive control, system adaptation, adaptive workload allocation, adaptive aiding, adaptive cruise control, flexible automation, adaptive mitigation strategies, human-systems inclusion, human-systems integration, multimodal interaction, biofeedback, augmented cognition, psychological adjusting, biofeedback, cognitive performance enhancement, cognitive enhancement, mental support, neurofeedback, cognitive	AI, artificial intelligence, digital intelligence, machine learning, analytics, big data, machine intelligence, deep learning, prediction model, neural network, computational intelligence, predictive models, affective computing, physiological computing, computer reasoning, decision support systems, big data, expert systems, statistical learning, artificial consciousness, machine consciousness (note of caution, very unspecific and badly worded), wearable computing, supervised learning, unsupervised learning, reinforcement learning, artificial neural network, artificial cognition, machine perception, attention-based computing, computer vision, support vector machines, natural language processing, convolutional neural network, convolutional network, computer heuristics, regression, transfer learning, genetic algorithm, recommender system, recommendation algorithm, random forest, linear classifier,

assistance system, emotional-based agent robotics, cyber-physical system, industrial robot, social robot, evolutionary robotics, cognitive robotics, aerial robotic, teleoperation, telerobotics, cyber-physical-human-system, human-cyber-physical system, human robot interaction, human machine interaction, collaborative robot, cobot, physical human-robot interaction, physical-robot-human interaction, shared robotic task, human-robot team, closed-loop human-robot interaction, real-time human-robot interaction, robotic symbiotic network, human-robot collaboration, safe physical human–robot collaboration, brain mediated human-robot interaction, admittance control, cognitive medical robot, smart health environment, medical robots and system, robot assisted surgery, robotic surgical procedures, computer-aided diagnosis, neuronavigation		monitoring	nearest neighbor search, decision tree, hidden markov chain, ensemble learning, relevance vector machine, brain-computer interface, brain-machine interface, neurotechnology

A1.4 Included manuscripts

Table 44

References of included manuscripts

Study ID	Reference	Title	Country	Publication type	Year
7577	(Toreini et al., 2020)	Using eye-tracking for visual attention feedback	Germany	Conference Proceeding	2020
7276	(Karthikeyan & Mehta, 2020)	Towards a Closed-Loop Neurostimulation Platform for Augmenting Operator Vigilance	United States	Conference Proceeding	2020
7067	(Wang et al., 2019)	The Effectiveness of EEG-Feedback on Attention in 3D Virtual Environment	China	Conference Proceeding	2019
6892	(Demazure et al., 2019)	Sustained attention in a monitoring task: Towards a neuroadaptative enterprise system interface	Canada	Conference Proceeding	2019
6320	(Peternel et al., 2018)	Robot adaptation to human physical fatigue in human robot co-manipulation	Italy	Journal article	2018
5701	(Szafir & Mutlu, 2012)	Pay attention! Designing adaptive agents that monitor and improve user engagement	United States	Conference Proceeding	2013
5683	(Parnandi & Gutierrez-Osuna, 2021)	Partial Reinforcement in Game Biofeedback for Relaxation Training	United States	Journal article	2018
4782	(Yuksel et al., 2016)	Learn piano with BACh: An adaptive learning interface that adjusts task difficulty based on brain state	United States	Conference Proceeding	2016
4250	(Schiatti et al., 2018)	Human in the Loop of Robot Learning: EEG-Based Reward Signal for Target Identification and Reaching Task	Italy	Conference Proceeding	2018
4154	(Raaijmakers et al., 2013)	Heart Rate Variability and Skin Conductance Biofeedback: A Triple-Blind Randomized Controlled Study	Netherlands	Conference Proceeding	2013
3887	(Kim et al., 2020)	Flexible online adaptation of learning strategy using EEG-based reinforcement signals in	Germany	Conference Proceeding	2020

		real-world robotic applications			
3716	(Dey et al., 2019)	Exploration of an EEG-Based Cognitively Adaptive Training System in Virtual Reality	Australia	Conference Proceeding	2019
3575	(Chaouachi et al., 2015)	Adapting to learners' mental states using a physiological computing approach	Canada	Journal article	2019
3501	(Causse et al., 2019)	Encoding decisions and expertise in the operator's eyes: Using eye-tracking as input for system adaptation	France	Journal article	2019
3438	(Ghandi et al., 2021)	Embodied empathy: Using affective computing to incarnate human emotion and cognition in architecture	United States	Journal article	2019
3379	(Larradet et al., 2017)	Effects of galvanic skin response feedback on user experience in gaze-controlled gaming: A pilot study	Italy	Conference Proceeding	2017
3293	(Zhou et al., 2015)	Dynamic workload adjustments in human-machine systems based on GSR features	Australia	Conference Proceeding	2015
3292	(Labonte-Lemoyne et al., 2018)	Dynamic threshold selection for a biocybernetic loop in an adaptive video game context	Canada	Journal article	2018
3277	(Breslow et al., 2014)	Dynamic operator overload: A model for predicting workload during supervisory control	United States	Journal article	2014
2339	(Azgomi et al., 2021)	Closed-Loop Cognitive Stress Regulation Using Fuzzy Control in Wearable-Machine Interface Architectures	United States	Journal article	2021
2218	(Ramos et al., 2021)	Building a Drone Operator Digital Twin using a Brain-Computer Interface for Emotion Recognition	Portugal	Conference Proceeding	2021
2201	(Di Flumeri et al., 2019)	Brain-Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems	Italy	Journal article	2019
2199	(Trachel et	Brain-computer interaction for online enhancement of	France	Journal	2018

	al., 2018)	visuospatial attention performance		article	
2187	(Tseng et al., 2012)	Brain Computer Interface-based Multimedia Controller	Taiwan	Conference Proceeding	2012
2149	(Pavlidis et al., 2021)	Biofeedback Arrests Sympathetic and Behavioral Effects in Distracted Driving	United States	Journal article	2021
1898	(Darzi & Novak, 2021)	Automated affect classification and task difficulty adaptation in a competitive scenario based on physiological linkage: An exploratory study	United States	Journal article	2021
1844	(Vortmann & Putze, 2020)	Attention-aware brain computer interface to avoid distractions in augmented reality	Germany	Conference Proceeding	2020
1606	(Nalepa et al., 2019)	Analysis and Use of the Emotional Context with Wearable Devices for Games and Intelligent Assistants	Poland	Journal article	2019
1580	(Zargari Marandi et al., 2019)	An oculometrics-based biofeedback system to impede fatigue development during computer work: A proof-of-concept study	Denmark	Journal article	2019
1314	(Zhang et al., 2021)	An Adaptive Attention Regulation Method Based on Biocybernetic Loop	China	Conference Proceeding	2020
1228	(Aranyi et al., 2016)	Affective Interaction with a Virtual Character Through an fNIRS Brain-Computer Interface	UK	Journal article	2016
1226	(Govindarajan et al., 2018)	Affective Driver State Monitoring for Personalized, Adaptive ADAS	United States	Conference Proceeding	2018
1000	(Lim et al., 2021)	Adaptive human-robot interactions for multiple unmanned aerial vehicles	Australia	Journal article	2021
926	(Arico, Borghini, Di Flumeri, Colosimo, Bonelli, et al., 2016)	Adaptive automation triggered by EEG-based mental workload index: A passive Brain-Computer Interface application in realistic air traffic control environment	Italy	Journal article	2016

555	(Arico, Borghini, Di Flumeri, Colosimo, Pozzi, et al., 2016)	A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks	Italy	Journal article	2016
471	(El-Samahy et al., 2015)	A new computer control system for mental stress management using fuzzy logic	UK	Journal article	2015

A1.5 Supplementary results

A1.5.1 Problem Space – Environment

Table 45

Domain and references of manuscripts

Domain	Count	%	Study ID
Aeronautics	7	19.44%	555, 926, 1000, 2201, 2218, 3277, 3501
Art	2	5.56%	2187, 3438
Business	4	11.11%	1580, 3293, 6892, 7577
Education	3	8.33%	3575, 4782, 5701
Robotics	3	8.33%	3887, 4250, 6320
Smart-Home	1	2.78%	1844
Training	7	19.44%	471, 1314, 2339, 3379, 3716, 4154, 5683
Transports	2	5.56%	1226, 2149
Video games	2	5.56%	1606, 3292
Virtual Agent Design	1	2.78%	1228
Unclear/Inferred	4	11.11%	1898, 2199, 7067, 7276

Table 46

Target tasks and references of manuscripts

Target Task	Count	%	Study ID
Cognitive Task	1	2.78%	3293
Computer Task	1	2.78%	1580
Concentration Task	1	2.78%	7067
Cooperation Task	1	2.78%	1898
Driving Task	2	5.56%	1226, 2149
Game Task	2	5.56%	1606, 3292
Human-Robot Task	3	8.33%	3887, 4250, 6320
Information Seeking Task	1	2.78%	7577
Learning Task	3	8.33%	3575, 4782, 5701
Listening Task	1	2.78%	2187
Natural Task	1	2.78%	3438

Relaxation Task	1	2.78%	5683
Self-Regulation Task	2	5.56%	3379, 4154
Supervision Task	2	5.56%	1000, 3277
Surveillance Task	7	19.44%	555, 926, 2201, 2218, 3501, 6892, 7276
Training Task	1	2.78%	3716
Unclear/Inferred	6	16.67%	471, 1228, 1314, 1844, 2199, 2339

Table 47

Target users and references of manuscripts

Target Users	Count	%	Study ID
Air traffic controllers	4	11.11%	555, 926, 2201, 3501
Drivers	2	5.56%	1226, 2149
Learners	4	11.11%	3575, 3716, 4782, 5701
Office workers	4	11.11%	1580, 2339, 6892, 7577
Operators	1	2.78%	7276
Production line workers	1	2.78%	6320
Unmanned Vehicule operator	3	8.33%	1000, 2218, 3277
Video game players	2	5.56%	1606, 3379
Unclear/Inferred	15	41.67%	471, 1228, 1314, 1844, 1898, 2187, 2199, 3292, 3293, 3438, 3887, 4154, 4250, 5683, 7067

A1.5.2 Problem Space – Problematization

Table 48

Research questions and objectives

	Research Question		Research Objective	
	Count	%	Count	%
Clearly reported	3	8.33%	21	58.33%
Unclear/inferred	33	91.67%	15	41.67%

Table 49

Research goal and research questions

		Research goal					
		Clearly Reported		Unclear/inferred			
		Count	%	Count	%	Total Count	Total %
Research Question	Clearly reported	3	0.083		0	3	0.083
	Unclear/inferred	18	0.5	15	0.417	33	0.917
		21	0.583	15	0.417	36	

Table 50

Artifact objectives

	Defined Artifact Objectives	Count	%
Clearly reported		22	0.611
	General objective	13	0.361
	Hypotheses	4	0.111
	Set of requirements	5	0.139
Unclear/inferred		14	0.389
	Research goal	3	0.083
	Unclear/inferred	11	0.306

A1.5.3 Solution Space - Target Users and Actual Subjects Comparison

Table 51

Target users and manuscripts participants

Study ID	Title	Target Users L2	Subjects L1
7577	Using eye-tracking for visual attention feedback	Office workers	Students
7276	Towards a Closed-Loop Neurostimulation Platform for Augmenting Operator Vigilance	Operators	Unclear/inferred
7067	The Effectiveness of EEG-Feedback on Attention in 3D Virtual Environment	Unspecified	Students
6892	Sustained attention in a monitoring task: Towards a neuroadaptative enterprise system interface	Office workers	Students
6320	Robot adaptation to human physical fatigue in human robot co-manipulation	Production line workers	Unclear/inferred
5701	Pay attention! Designing adaptive agents that monitor and improve user engagement	Learners	Students
5683	Partial Reinforcement in Game Biofeedback for Relaxation Training	Unspecified	Students
4782	Learn piano with BACh: An adaptive learning interface that adjusts task difficulty based on brain state	Learners	Unclear/inferred
4250	Human in the Loop of Robot Learning: EEG-Based Reward Signal for Target Identification and Reaching Task	Unclear/inferred	Unclear/inferred
4154	Heart Rate Variability and Skin Conductance Biofeedback: A Triple-Blind Randomized Controlled Study	Unclear/inferred	Students
3887	Flexible online adaptation of learning strategy using EEG-based reinforcement signals in real-world robotic applications	Unclear/inferred	Unclear/inferred
3716	Exploration of an EEG-Based Cognitively Adaptive Training System in Virtual Reality	Learners	Students
3575	Adapting to learners' mental states using a physiological computing approach	Learners	Students
3501	Encoding decisions and expertise in the operator's eyes: Using eye-tracking as input for system adaptation	Air traffic controllers	Aeronautic Students
3438	Embodied empathy: Using affective computing to incarnate human emotion and cognition in architecture	Unclear/inferred	Unclear/inferred

3379	Effects of galvanic skin response feedback on user experience in gaze-controlled gaming: A pilot study	Video game players	Unclear/inferred
3293	Dynamic workload adjustments in human-machine systems based on GSR features	Unclear/inferred	Students
3292	Dynamic threshold selection for a biocybernetic loop in an adaptive video game context	Unclear/inferred	Students
3277	Dynamic operator overload: A model for predicting workload during supervisory control	Unmanned Vehicule operator	Psychology Students
2339	Closed-Loop Cognitive Stress Regulation Using Fuzzy Control in Wearable-Machine Interface Architectures	Office workers	Unclear/inferred
2218	Building a Drone Operator Digital Twin using a Brain-Computer Interface for Emotion Recognition	Unmanned Vehicule operator	Unclear/inferred
2201	Brain-Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems	Air traffic controllers	Air Traffic Controllers
2199	Brain-computer interaction for online enhancement of visuospatial attention performance	Unclear/inferred	Unclear/inferred
2187	Brain Computer Interface-based Multimedia Controller	Unclear/inferred	Unclear/inferred
2149	Biofeedback Arrests Sympathetic and Behavioral Effects in Distracted Driving	Drivers	Students
1898	Automated affect classification and task difficulty adaptation in a competitive scenario based on physiological linkage: An exploratory study	Unclear/inferred	Students
1844	Attention-aware brain computer interface to avoid distractions in augmented reality	Unclear/inferred	Students
1606	Analysis and Use of the Emotional Context with Wearable Devices for Games and Intelligent Assistants	Video game players	Students
1580	An oculometrics-based biofeedback system to impede fatigue development during computer work: A proof-of-concept study	Office workers	Unclear/inferred
1314	An Adaptive Attention Regulation Method Based on Biocybernetic Loop	Unclear/inferred	Unclear/inferred
1228	Affective Interaction with a Virtual Character Through an fNIRS Brain-Computer Interface	Unclear/inferred	Unclear/inferred
1226	Affective Driver State Monitoring for Personalized, Adaptive ADAS	Drivers	Unclear/inferred

1000	Adaptive human-robot interactions for multiple unmanned aerial vehicles	Unmanned Vehicule operator	Aeronautic Students
926	Adaptive automation triggered by EEG-based mental workload index: A passive Brain-Computer Interface application in realistic air traffic control environment	Air traffic controllers	Air Traffic Controllers
555	A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks	Air traffic controllers	Air Traffic Controllers
471	A new computer control system for mental stress management using fuzzy logic	Unclear/inferr ed	Unclear/infer red

A2 Chapter 3

A2.1 Literature review query

Example of query use in Web of Science

```
TS=("deep learning" OR "neural network") AND TS=("working memory" OR "workload" OR "mental workload" OR "cognitive load" OR "cognitive workload") AND TS=("electroencephalography" OR "EEG")
```

A2.2 Literature review

Table 52

Literature review table

Study – Title (Authors)	Instrument	Problem Setting	Task (condition)	Inputs Domain	Data curation	Features	Model Name	Layer	Activation Function	Optimizers	Regularization	Training Strategies	Design choices	Baseline	Accuracy
EEG-Signals Based Cognitive Workload Detection of Vehicle Driver using Deep Learning (Mohammad A. Almogbel et al., 2018)	EEG	Driver workload level, with subject	Driving Simulator (dense traffic vs low traffic), within subject	Time, raw	No	**End-to-end**	CNN	7	Hidden layers: ReLu Output Layers: Softmax	RMSProp (lr =0.002)	Normalized by z-score	Overlapping windows at different time	Windows size # Layers	Deep learning with hand engineered features (1L CNN)	95.31%
Cognitive Workload Detection from Raw EEG-Signals of Vehicle Driver using Deep Learning (Almogbel, Dang, Kameyama, et al., 2019)	EEG	Driver workload level, within subject,	Driving Simulator (dense traffic vs low traffic vs zero traffic), within subject	Time, raw	No	**End-to-end**	CNN	8	Hidden layers: ReLu Output Layers: Softmax	RMSProp (lr=0.0001)	Normalized by z-score	Overlapping windows at different time	Windows size	Deep learning with hand engineered features (1L CNN)	96%
Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification (Baldwin & Penaranda, 2012)	EEG	Task Workload Classification, within and cross task, within subject	Reading span (2 level), Visuospatial n-back (2 level), Sternberg task (2 level)	Frequency, values	band-pass: 0.1 - 70 Hz Temporary artifacts Noisy channels	8th order elliptical filters with stopband attenuation of 20 dB and passband ripple of 1 dB: delta (0.01–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (13–30 Hz), and gamma (31–42 Hz) / per channels	Multi-layer perceptrons					5 seconds non-overlapping windows			Within = 87.1% Cross = 44.8%
Artificial Neural Network classification of operator workload with an assessment of time variation and noise-enhancement to increase performance (Casson, 2014)	EEG	Within, time effect post training, cross day	Flight simulator task (low and high)	Frequency, values		FFT Frequency bands: 0–4 Hz, 4–7 Hz, 7–12 Hz, 12–30 Hz, 30–42 Hz, 42–84 Hz, 84–128 Hz	ANN	5			Features are zero mean and unit standard deviation normalized	30 seconds windows, 25 sec overlapping Trained on the first 20 epochs, remaining epochs (11) used for testing	Time independent average performance Temporal performance Temporal performance with noise		73% average

													enhancement techniques		
Identification of Mental Workload Using Imbalanced EEG Data and DySMOTE-based Neural Network Approach (Cui et al., 2016)	EEG + ECG	Mental workload, unbalanced dataset, Within subject	Automation-enhanced Cabin Air Management System (ACAMS) with complex tasks (Auto vs Manual (5 levels))	Frequency, values	Blink artifact Butterworth IIR filter band-pass 0-40 Hz	Butterworth IIR filter delta (1-4Hz), theta(5-8Hz), alpha(9-12Hz), beta(13-32Hz), gamma (33-40Hz)	DySMOTE					5 fold cross validation (5 iterations per model)	Sampling scheme: SMOTE oversampling technique (vs CSUS, CSOS, DyROS)	CSUS, CSOS, DyROS	50 % to 70 %
Classification of Movement Direction From Electroencephalogram During Working Memory Time (Fukuda et al., 2019)	EEG	Motor movement during working memory task, within subject		Frequency, values	Trial rejection Blink artifact BTW (8th order) low cut-ff 49 Hz	FFT for 1-40 Hz, Phase (with imaginary parts)	Neural Network	3	Hidden: Sigmoid		Sparse regularization of L1 L1 norm Dropout (0.5)	5 fold cross validation			62%
EEG-Based Spatio-Temporal Convolutional Neural Network for Driver Fatigue Evaluation (Gao et al., 2019)	EEG	Driver fatigue, within-subject	Driving simulator. Highway	Time, raw	(1000 Hz, downsample 100 Hz), 1-50 Hz	**End-to-end**	EEG-based spatial–temporal convolutional neural network (ESTCNN)	14	Hidden: ReLu Output: Softmax	SGD (lr = 0.001)	Batch normalization	10 fold cross validation	Core block: three convolutional blocks and a pooling layer.	SVM, LSTM, CNNs, FFT + CNNs	97.37%
Cross-Participant EEG-Based Assessment of Cognitive Workload Using Multi-Path Convolutional Recurrent Neural Networks (Hefron, Borghetti, Kabban, et al., 2018)	EEG	Cognitive Workload, Cross subject	low-workload MATB, verbal N-back (4 level)	Frequency, values	High pass 1 Hz Interpolate bad channels Line noise High variance artifact	PSD (3 and 55 Hz) / electrodes, STFT 2s hanning window, 1 sec overlap	MPCRNN		ReLu	Adam	L1/L2 regularization batch normalization dropout (0.2)	Zero-data cross participant, cross validated group (7 fold), validation se group	Individually trained models vs group-trained models Temporal sequence length Ensemble	Two NN (single-hidden-layer ANN, 2 layers LSTM)	86.80%
Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation (Hefron et al., 2017)	EEG	Cognitive Workload, Within-subject, cross day	Multi-Attribute Task Battery (MATB)	Frequency, values		frequency bands (delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (15–30), and gamma (30–55 Hz) Power spectral density (Morlet) Mean, variance, skewness, and kurtosis of the power distribution	SVM Radial basis function ANN RNN LSTM		Sigmoid	Adam	Dropout	4 fold cross validation	Models		93.00%
Novel functional brain network methods based on CNN with an application in proficiency evaluation (Hua et al., 2019)	EEG	Working memory, within subject	Visuo-spatial working memory task	Time + Frequency, raw + values + images	frequency bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (15–30 Hz), low gamma (30–45 Hz) and all (0.5–45	**End-to-end + FIR filters for rhythm extraction**	brain connection based on CNN (BCCNN)						Type of inputs (fusions)		99%, 96.35%

					Hz)										
Deep Convolutional Neural Networks for mental load classification based on EEG data (Jiao et al., 2018)	EEG	Mental load, within subject	Working memory, letters	Frequency, time, image		FFT: theta (4–7 Hz), alpha (8–13 Hz), and beta (13–30 Hz) (sum of squared absolute values)	CNNs	7	ReLu Softmax	SGD	Dropout (0.5)	SVM, CNNs, CNN+LSTM, CNN + 1D Conv	Single-model v double-model method Input data (2-D scalp power maps / temporal information) PGBM	SVM, LSTM, CNNs, FFT + CNNs	
Across-subject estimation of 3-back task performance using EEG signals (Kim et al., 2014)	EEG	Working Memory,	N-Back, letters (0-3)	Frequency, values	Band-passed 0.1 - 40 Hz, re-ref to common average, baseline correction	STFT (-0.5:2s)	MLP, ANN					leave-one-out cross-subject	Ensembles		
Investigating Ensemble Learning and Classifier Generalization in a Hybrid, Passive Brain-Computer Interface for Assessing Cognitive Workload (Klosterman & Eepp, 2019)	EEG, HEOG, VEOG, ECG	mental Workload, Within-subject, cross day	Multi-Attribute Task Battery (MATB)	Fusion: Frequency (EEG), Engineered Features for ECG and VEOG	Fz, F7, Pz, T5, and O2. HEOG and VEOG artifact using a post-hoc regression method Blink Detection	Inter-Beat interval for ECG delta (0.5–3 [Hz]), theta (4–7 [Hz]), alpha (8–12 [Hz]), beta (13–30 [Hz]) and gamma (31–42 [Hz]). Blink Rate	ANN, LIN-SVM, LDA	3				k fold cross-validation Ensemble tested within and across day	Ensembles, AdaBoost Number of day in the dataset		
Cognitive Analysis Of Working Memory Load From Eeg, By A Deep Recurrent Neural Network (Kuanar, Athitsos, Pradhan, Mishra, Rao, et al., 2018)	EEG	Cognitive Load, cross subject	Working memory (n0back), 4 level	Image, Frequency + 2D projection of sensors	Bandpass	FFT: theta (4-7Hz), alpha (8-13Hz), beta (13-30 Hz).	RNN + ConvNet	12	hidden: ReLu Output: Softmaz	SGD Adam	L2 Regularization Dropout (0.5) Gaussian Noise	leave subject-out cross validation technique	Hybrid models (ConvNet, LSTM, Bidirectional LSTM)	Random Forest (RF), Support Vector Machines, Logistic Regression	92.50%
Classification of Working Memory Performance from EEG with Deep Artificial Neural Networks (Kwak et al., 2019)	EEG	Performance discrimination based on working memory, between	Working memory, letters	Frequency, power ratios		FFT: theta (4-8 Hz), alpha (8-14 Hz), beta (14-30Hz), gamma1 (30-50 Hz) and gamma2 (50-100Hz).	ANN	5	Relu Softmax	Adam		leave-one-subject-out cross validation			61%
Convolutional Neural Networks with Large-Margin Softmax Loss Function for Cognitive Load Recognition (Liu & Liu, 2017)	EEG	Cognitive load recognition	Working memory, letters (4 level)	Frequency, values + images		FFT: theta (4-7 Hz), alpha (8-13 Hz) and beta (13-30 Hz), sum of squared absolute values within every band for every electrode	ConvNets	6	PReLU	SGD	Weight decay strategy batch normalization		1D-2D Conv nets + large-margin SoftMax loss functions + Batch size		93%, 92%
Assessment of Cognitive Load using Multimedia Learning and	EEG	Cognitive load recognition	Resting and learning tasks	Time-Frequency, Image	Bandpass: 0-100HZ Eyes-artifact,	Discrete Wavelet Transform: 0-4 Hz for delta, 3-7	CNN, ResNet					Transfer Learning	Network Complexity		

Resting States with Deep Learning Perspective (Qayyum, Faye, et al., 2018)					heart rate, muscle movement	for theta, 7-12 for alpha and 15-28 for beta wave.									
Classification of EEG Learning and Resting States using 1D-Convolutional Neural Network for Cognitive Load Assessment (Qayyum, Khan, et al., 2018)	EEG	Cognitive load recognition	Resting and learning tasks	Time-Frequency, values	Z-score normalization, Eyes-artifact, heart rate, muscle movement	Discret Wavelet Transform: 0-4 Hz for delta, 3-7 for theta, 7-12 for alpha and 15-28 for beta wave.	1D-CNN	7	ReLu	SGD	Batch normalization		Choice of features		Between 85.6% and 93.2
Ternary-task convolutional bidirectional neural turing machine for assessment of EEG-based cognitive workload (Qiao & Bi, 2020)	EEG	Cognitive workload estimation, between subject	Sternberg memory experiment (4 levels)	Frequency, Images		Azimuthal Equidistant Projection + FFT (three frequency bands(0–7, 7–14, 14–49 Hertz).	Ternary-task Convolutional Bidirectional Neural Turing Machine (TT-CBNTM) convolutional neural network (CNN) + bidirectional neural turing machine (BNTM).		ReLu	SGD RMSProp optimizer	Ternary-task regularization framework L2 norm Dropout Early-stopping	10-fold cross-validation	VGG architecture, dropout parameters, kernel size	SVM, DBN-3,VGGs, BNTM-3, NTM-3, C-SAE, C-RNN, , C-GRU, C-NTM, ConvNet + LSTM/1D-Conv	96.30%
Multimodal fNIRS-EEG Classification Using Deep Learning Algorithms for Brain-Computer Interfaces Purposes (Saadati et al., 2020b)	FNIRS + EEG	Mental workload classification, within subject	n-back (4 levels), discrimination/selection response task (DSR), word generation (WG) tasks	Time, values	bandpass 1-40 Hz	Event-related desynchronization and synchronization analysis	DNN	5	Elu and Relu				Multiple window lengths	SVM	87%
Convolutional Neural Network for Hybrid fNIRS-EEG Mental Workload Classification (Saadati et al., 2020a)	FNIRS + EEG	Mental workload classification, within subject	N-back (4 levels)	Time, values	band pass: 1-40hz	Event-related desynchronization and synchronization analysis	CNNs		Elu and Relu		Dropout		Windows size, Activation function	SVM	89%
EEG-Based Mental Workload Estimation (Samima & Sarma, 2019)	EEG	Mental workload Estimation, within subject	Working Memory Test Battery (verbal and visuo-spatial) (3 levels)	Frequency, values + ratios	Bandpass: 0.5-45Hz Notch: 50Hz Normalization Artifact removal	EEG rhythms namely, (4.0 - 7.9Hz); (7.9 - 10.0Hz); (10.0 - 13:0Hz); (13.0 - 18.0Hz); (18.0 - 25.0Hz); (25.0 - 45.0Hz) Spectral powers	ANN	1	tansig				Complete trial windows size		96.6
Cognitive Load Recognition Using Multi-channel Complex Network	EEG	Cognitive load recognition		Frequency, Images		FFT: theta (4–7 Hz), alpha (8–12 Hz) and beta (13–30 Hz)	Multi-channel network + SVM					10-fold cross validation and test		ConvNet + 1D-Conv ConvNet + LSTM	33.86%, 32.28%, 85.66%, 78.65%

Method (Shang et al., 2017)						Network Structural Features						error		ConvNet + LSTM / 1D-Conv	and 86.33%
Individual-Specific Classification of Mental Workload Levels Via an Ensemble Heterogeneous Extreme Learning Machine for EEG Modeling (Tao et al., 2019)	EEG	Mental Workload, Within-subject and across	HM collaboration environment, Automation-enhanced cabin air management systems (ACAMS)	Frequency, values	Filtered via ICA	FFT: (4–8 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31-40) power difference between hemispheres Mean, variance, zero, crossing rate, Shannon entropy, spectral entropy, kurtosis, and skewness	extreme learning machine (ELM) heterogeneous ensemble ELM (HE-ELM) Adaboost	1	hardlim, sigmoid, sine	Gradient Descent			Activation function, Depth	K-nearest neighbor (KNN), artificial neural network with single hidden layer (ANN), denoising autoencoder (DAE), logistic regression (LR), Adaboost based on the decision tree, stacked denoising autoencoder (SDAE), LPP-KNN, LPP-ANN, LPP-DAE, LPP-LR, LPP-AD, LPP-SDAE, and LPP-NB.	0.88
Cross-subject workload classification with a hierarchical Bayes model (Wang et al., 2012)	EEG	workload, cross-subject	MATB - Multi-Attribute Task Battery (3 level)	Frequency, values	Down-sampled to 128 Hz Bandpass 0.05 Hz - 100 Hz	STFT (40 sec windows, 35 sec overlap): delta [2–4 Hz], theta [5–8 Hz], alpha [9–13 Hz], beta [14–32 Hz] and gamma [33–43 Hz]), gamma bands ([33–57 Hz] and [63–100 Hz])	Hierarchical Bayes model					fivefold cross-validation	# of hidden stats in NB	1 Layer NN (trained within subject)	around 80%
Pilots' Fatigue Status Recognition Using Deep Contractive Autoencoder Network (Wu et al., 2019)	EEG	Fatigue, within subject	Flight Simulator, take, approach and landing	Frequency, ratios	Downsampled at 160 Hz.	Wavelet packet transform: delta, theta, alpha and beta, channel: Fp1	deep contractive autoencoder network	3	Softmax			fivefold cross-validation	window functions		91.67%
Assessing cognitive mental workload	EEG	Mental Workload,	automatic enhanced cabin air	Frequency, values	Butterworth filter with a	FFT for power spectral density:	Ensemble Stacked	2				10-fold cross	Number of hidden	LR, NB, ELM,	0.8525

via EEG signals and an ensemble deep learning classifier based on denoising autoencoders (Yang et al., 2019)		within subject	management system (ACAMS)		lowpass cut-off frequency of 40 Hz Independent component analysis was employed to correct the ocular artifacts	average power in the theta (4–8 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–40 Hz) bands (2sec windows) EEG temporal features including mean, variance, zero crossing rate, Shannon entropy, spectral entropy, kurtosis, and skewness	Denoising Autoencoder (Ensemble SDAE)					validation	layers Number of hidden neurons Input: frequency vs time	KNN	
Recognition of Cognitive Task Load Levels Using Single Channel EEG and Stacked Denoising Autoencoder (Yin & Zhang, 2016)	EEG	Cognitive Task Load, within subject	automatic enhanced cabin air management system (ACAMS)	Frequency, values	Butterworth filter, 4th order, Band pass: 1.5-40 Hz Eye blink correction via ICA	FFT, Single channel	Stacked Denoising Autoencoder	7					Channel Selection		74%
Cross-subject recognition of operator functional states via EEG and switching deep belief networks with adaptive weights (Yin & Zhang, 2017b)	EEG	Operator Functional States, cross subject	AutoCAMS	Frequency and more, values	Butterworth IIR Filter, ICA for blink correction	frequency, log-energy entropy, mean, five power components, Shannon entropy, sum of energy, variance, zero-crossing rate of each channel and power differences between four channel pairs.	Switching Deep Belief Network							KNN, NB, LR, LSSVM, SAE, DBN, PCA–KNN, PCA–NB, PCA–LR, PCA–LSSVM, PCA–SAE, and PCA–DBN	62%, 71%, and 40%
Cross-session classification of mental workload levels using EEG and an adaptive deep learning model (Yin & Zhang, 2017a)	EEG	Mental workload, within subject, cross session	AutoCAMS, 4 levels	Frequency, values	4-order Butterworth IIR filter: low-pass frequency of 40 Hz Eye blink correction via ICA	FFT: average power in theta (5–7.5 Hz), alpha(8–13.5 Hz), beta1 (14–20 Hz), beta2 (20.5–30 Hz),	Adaptive Stacked Denoising Autoencoder	6	Sigmoid	Gradient Descent			Adaptive Shallow layer Within Session, Cross session Channel selection, features selection Noise robustness	ANN, NB, KNN, SVM, BSVM	Within: 0.95 Cross session: 0.87
Task-generic mental fatigue recognition based on	EEG	Mental Fatigue, cross task	AutoCAMS, 2 designs, 4 levels	Frequency and more, values	4-order Butterworth IIR filter: bandpass .5 -	Average power theta, alpha, beta (13–30 Hz), and gamma	dynamical deep extreme learning	10	logistic sigmoid				Complexity	H-ELM, ELM, LSSVM, ANN, LR,	0.7551, 0.6551

neurophysiological signals and dynamical deep extreme learning machine (Yin & Zhang, 2018)					40 Hz Eye blink correction via ICA	(30–40 Hz) bands, mean, variance, zero-crossing rate (ZCR), Shannon entropy (denoted as Entropy 1), log-energy entropy (denoted as Entropy 2), kurtosis, and skewness	machine				NB, KNN, PCA-NB, PCA-KNN, PCA-LSSVM, PCA-NB, PCA-ELM.	
Physiological-signal-based mental workload estimation via transfer dynamical autoencoders in a deep learning framework (Yin et al., 2019)	EEG	Mental Workload, cross subject	Two process control tasks, one emotion stimuli	Frequency and more, values	downsampled 128 Hz.	average power spectral density (PSD) within frequency bands of theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz) and gamma (30–40 Hz), Band power via FFT, mean, variation, zero crossing rate, Shannon entropy, spectral entropy, kurtosis, and skewness across	transfer dynamical autoencoder	logistic sigmoid		Complexity	NB, LR, KNN, ANN, ELM, LSSVM, SAE, DBN, CNN, TDAE	0.8623, 0.8987

A2.3 Manipulation check

Figure 41

Normal quantile-quantile plot

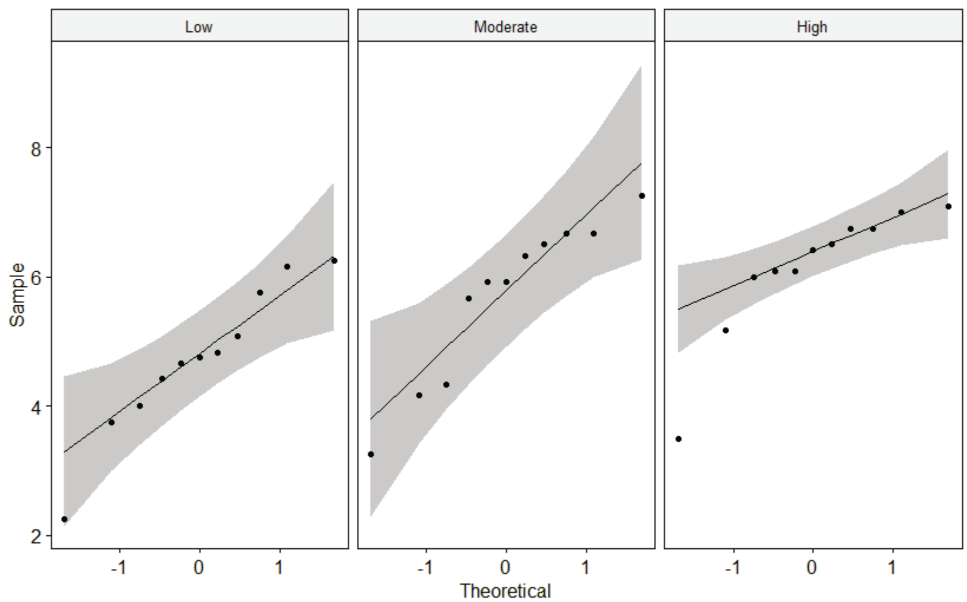


Table 53

Shapiro-Wilk test

Maneuver	variable	statistic	p
Low	Workload	0.945	0.576
Moderate	Workload	0.897	0.171
High	Workload	0.804	0.0107

Table 54

Mauchly's test for sphericity

Effect	W	P
Maneuver	0.59	0.093

Table 55

Greenhouse-Geisser sphericity corrections

Effect	GGe	DF[GG]	p[GG]	p[GG]<.05	HFe	DF[HF]	p[HF]	p[HF]<.05
Maneuver	0.709	1.42	14.18	0.008	0.793	1.59	15.85	0.006

Table 56

ANOVA

Effect	DFn	DFd	F	p	ges
Maneuver	2	20	7.995	0.003	0.224

Table 57

Pairwise T-Tests

.y.	group1	group2	n1	n2	statistic	df	p	p.adj	p.adj.signif
Workload	Low	Moderate	11	11	-4.53	10	0.001	0.003	*
Workload	Low	High	11	11	-3.40	10	0.007	0.02	*
Workload	Moderate	High	11	11	-1.03	10	0.329	0.987	ns

Note. Adjustment: Bonferroni

A2.4 Benchmark statistics

We use Wilcoxon signed-rank tests to compute the statistical significance of mean difference of performance metrics between deep learning models as suggested by (Benavoli, Corani, & Mangili, 2016). For multiple tests correction, we used Benjamini-Hochberg procedure for false-discovery-rate (10 %) correction with a nominal α = 0.05 (Benjamini & Hochberg, 1995).

Table 58

Wilcoxon signed rank test comparing the baselines

clf1	clf2	p_value	statistic	sig	critical_p_value
COV+RMDM	LR	0	0	TRUE	0.007
COV+LR	LR	0	0	TRUE	0.013
COV+LR	ppn	0	0	TRUE	0.02
COV+RMDM	ppn	0	0.5	TRUE	0.027
COV+RMDM	PCA+CSP+LDA	0	11	TRUE	0.033
COV+LR	PCA+CSP+LDA	0	17.5	TRUE	0.04
COV+RMDM	Xdawn+LR	0	38	TRUE	0.047
COV+LR	Xdawn+LR	0	71.5	TRUE	0.053
Xdawn+LR	ppn	0	73.5	TRUE	0.06
LR	Xdawn+LR	0	117	TRUE	0.067
PCA+CSP+LDA	ppn	0	159.5	TRUE	0.073
LR	PCA+CSP+LDA	0	202	TRUE	0.08
PCA+CSP+LDA	Xdawn+LR	0.117	379	FALSE	0.087
COV+LR	COV+RMDM	0.313	428.5	FALSE	0.093
LR	ppn	0.504	458.5	FALSE	0.1

Table 59

Wilcoxon signed rank test comparing deep learning models

clf1	clf2	p_value	statistic	sig	critical_p_value
FCN	MLP	0	0	TRUE	0.002
CNN	FCN	0	0	TRUE	0.004
FCN	MCDCNN	0	0	TRUE	0.005
DeepConvNet	FCN	0	0	TRUE	0.007
Encoder	FCN	0	0	TRUE	0.009
MLP	ResNet	0	0	TRUE	0.011
DeepConvNet	ResNet	0	0	TRUE	0.013
CNN	COV+RMDM	0	0	TRUE	0.015
FCN	RNN_LSTM	0	0	TRUE	0.016
RNN_LSTM	ResNet	0	0	TRUE	0.018
COV+RMDM	MLP	0	0	TRUE	0.02
CNN	ResNet	0	0	TRUE	0.022
Encoder	ResNet	0	0	TRUE	0.024
COV+RMDM	DeepConvNet	0	0	TRUE	0.025
EEGNet	RNN_LSTM	0	0	TRUE	0.027
MCDCNN	ResNet	0	0.5	TRUE	0.029
COV+RMDM	RNN_LSTM	0	0.5	TRUE	0.031
RNN_LSTM	ShallowConvNet	0	0.5	TRUE	0.033
COV+RMDM	Encoder	0	1	TRUE	0.035
EEGNet	MCDCNN	0	1.5	TRUE	0.036
EEGNet	Encoder	0	2	TRUE	0.038
DeepConvNet	EEGNet	0	4.5	TRUE	0.04
COV+RMDM	MCDCNN	0	4.5	TRUE	0.042
CNN	EEGNet	0	5	TRUE	0.044
MCDCNN	ShallowConvNet	0	7.5	TRUE	0.045
MLP	ShallowConvNet	0	7.5	TRUE	0.047
EEGNet	MLP	0	8	TRUE	0.049
DeepConvNet	ShallowConvNet	0	8	TRUE	0.051
CNN	ShallowConvNet	0	9.5	TRUE	0.053
Encoder	ShallowConvNet	0	14	TRUE	0.055
FCN	ShallowConvNet	0	27.5	TRUE	0.056

ResNet	ShallowConvNet	0	56	TRUE	0.058
EEGNet	FCN	0	71.5	TRUE	0.06
COV+RMDM	FCN	0	135.5	TRUE	0.062
EEGNet	ResNet	0	140	TRUE	0.064
COV+RMDM	ResNet	0.001	220.5	TRUE	0.065
COV+RMDM	ShallowConvNet	0.006	277.5	TRUE	0.067
CNN	DeepConvNet	0.007	278	TRUE	0.069
EEGNet	ShallowConvNet	0.009	288	TRUE	0.071
FCN	ResNet	0.027	325.5	TRUE	0.073
DeepConvNet	Encoder	0.044	340	TRUE	0.075
CNN	MLP	0.077	361.5	FALSE	0.076
DeepConvNet	RNN_LSTM	0.171	396.5	FALSE	0.078
DeepConvNet	MLP	0.207	406	FALSE	0.08
DeepConvNet	MCDCNN	0.276	421.5	FALSE	0.082
Encoder	MLP	0.33	431.5	FALSE	0.084
MLP	RNN_LSTM	0.336	432.5	FALSE	0.085
CNN	MCDCNN	0.415	445.5	FALSE	0.087
CNN	Encoder	0.458	452	FALSE	0.089
COV+RMDM	EEGNet	0.461	452.5	FALSE	0.091
Encoder	MCDCNN	0.465	453	FALSE	0.093
MCDCNN	RNN_LSTM	0.5	458	FALSE	0.095
CNN	RNN_LSTM	0.501	458	FALSE	0.096
MCDCNN	MLP	0.671	480	FALSE	0.098
Encoder	RNN_LSTM	0.734	487.5	FALSE	0.1

Table 60

Wilcoxon signed rank test comparing the optimizers

clf1	clf2	p_value	statistic	sig	critical_p_value
fcn_nadam	resnet_sgd	2.00E-05	148	TRUE	0.00357
fcn_adadelta	resnet_sgd	0.00018	190.5	TRUE	0.00714
fcn_sgd	resnet_sgd	0.00018	189.5	TRUE	0.01071
fcn_nadam	resnet_nadam	0.00347	261.5	TRUE	0.01429
fcn_adadelta	resnet_nadam	0.00652	280	TRUE	0.01786
fcn_adam	resnet_sgd	0.00661	280.5	TRUE	0.02143
fcn_sgd	resnet_nadam	0.00857	287.5	TRUE	0.025
resnet_adam	resnet_sgd	0.00883	288	TRUE	0.02857
fcn_nadam	resnet_adadelta	0.01875	312.5	TRUE	0.03214
fcn_adam	resnet_nadam	0.02847	326	TRUE	0.03571
fcn_adadelta	resnet_adadelta	0.03703	336	TRUE	0.03929
fcn_sgd	resnet_adadelta	0.06387	355.5	FALSE	0.04286
resnet_adam	resnet_nadam	0.07256	361	FALSE	0.04643
resnet_adadelta	resnet_sgd	0.08938	369	FALSE	0.05
resnet_adadelta	resnet_nadam	0.14461	390.5	FALSE	0.05357
fcn_adam	fcn_nadam	0.16645	397.5	FALSE	0.05714
fcn_nadam	resnet_adam	0.18904	403	FALSE	0.06071
fcn_adam	fcn_sgd	0.22723	412.5	FALSE	0.06429
fcn_sgd	resnet_adam	0.24738	416.5	FALSE	0.06786
fcn_adam	resnet_adadelta	0.25018	417.5	FALSE	0.07143
fcn_adadelta	resnet_adam	0.34362	435	FALSE	0.075
resnet_nadam	resnet_sgd	0.50394	459	FALSE	0.07857
fcn_adam	resnet_adam	0.58217	469.5	FALSE	0.08214
fcn_adadelta	fcn_sgd	0.59073	471	FALSE	0.08571
fcn_adadelta	fcn_nadam	0.69892	484	FALSE	0.08929
fcn_adadelta	fcn_adam	0.72187	486.5	FALSE	0.09286
fcn_nadam	fcn_sgd	0.96781	514	FALSE	0.09643
resnet_adadelta	resnet_adam	0.96807	514	FALSE	0.1

Table 61

Wilcoxon signed rank test comparing the dropout rate for FCN

clf1	clf2	p_value	statistic	sig	critical_p_value
FCN_.1	FCN_.4	0	90.5	TRUE	0.017
FCN_.3	FCN_.4	0	93.5	TRUE	0.033
FCN_.2	FCN_.4	0	103	TRUE	0.05
FCN_.1	FCN_.3	0.009	290.5	TRUE	0.067
FCN_.1	FCN_.2	0.014	305	TRUE	0.083
FCN_.2	FCN_.3	0.593	471	FALSE	0.1

Table 62

Wilcoxon signed rank test comparing the dropout rate for ResNet

clf1	clf2	p_value	statistic	sig	critical_p_value
ResNet_.3	ResNet_.4	0.206	407	FALSE	0.017
ResNet_.1	ResNet_.4	0.209	407.5	FALSE	0.033
ResNet_.2	ResNet_.4	0.221	410.5	FALSE	0.05
ResNet_.1	ResNet_.3	0.475	455	FALSE	0.067
ResNet_.2	ResNet_.3	0.484	456.5	FALSE	0.083
ResNet_.1	ResNet_.2	0.796	495	FALSE	0.1

Table 63

Wilcoxon signed rank test comparing the activation functions

clf1	clf2	p_value	statistic	sig	critical_p_value
fcn_elu	resnet_elu	0.002	244.5	TRUE	0.017
fcn_elu	fcn_relu	0.002	247.5	TRUE	0.033
resnet_elu	resnet_relu	0.017	309	TRUE	0.05
fcn_relu	resnet_relu	0.079	364.5	FALSE	0.067
fcn_elu	resnet_relu	0.692	483	FALSE	0.083
fcn_relu	resnet_elu	0.806	496	FALSE	0.1

Table 64

Wilcoxon signed rank test comparing the window size

clf1	clf2	p_value	statistic	sig	critical_p_value
resnet_1.5s	resnet_3s	0	102.5	TRUE	0.017
fcn_3s	resnet_1.5s	0	147	TRUE	0.033
fcn_1.5s	resnet_1.5s	0	186	TRUE	0.05
fcn_1.5s	resnet_3s	0.004	263.5	TRUE	0.067
fcn_1.5s	fcn_3s	0.017	307	TRUE	0.083
fcn_3s	resnet_3s	0.806	496	FALSE	0.1

Table 65

Wilcoxon signed rank test comparing FCN, ResNet, and baselines on the replication dataset

clf1	clf2	p_value	statistic	sig	critical_p_value
FCN	LR	0	1.5	TRUE	0.005
FCN	PCA+CSP+LDA	0	6.5	TRUE	0.01
LR	ResNet	0	22.5	TRUE	0.014
PCA+CSP+LDA	ResNet	0	24.5	TRUE	0.019
COV+RMDM	LR	0	45.5	TRUE	0.024
FCN	Xdawn+LR	0	48.5	TRUE	0.029
COV+LR	LR	0	54	TRUE	0.033
COV+RMDM	PCA+CSP+LDA	0	67.5	TRUE	0.038
COV+LR	PCA+CSP+LDA	0	80.5	TRUE	0.043
ResNet	Xdawn+LR	0	153.5	TRUE	0.048
COV+LR	Xdawn+LR	0	309	TRUE	0.052
COV+RMDM	Xdawn+LR	0	354	TRUE	0.057
LR	Xdawn+LR	0	375	TRUE	0.062
COV+LR	FCN	0	582.5	TRUE	0.067
PCA+CSP+LDA	Xdawn+LR	0	691	TRUE	0.071
COV+RMDM	FCN	0	692	TRUE	0.076
COV+RMDM	ResNet	0	753.5	TRUE	0.081
COV+LR	ResNet	0.001	772.5	TRUE	0.086
LR	PCA+CSP+LDA	0.163	1161	FALSE	0.09
FCN	ResNet	0.793	1375.5	FALSE	0.095
COV+LR	COV+RMDM	0.816	1381	FALSE	0.1

A2.5 Validation statistics

Table 66

Intra-model agreement descriptive statistics

Subject	Kappa Score		CCC		MCC	
	Mean	Std	Mean	Std	Mean	Std
1	0.663	0.085	0.76	0.088	0.675	0.083
2	0.116	0.136	0.131	0.112	0.163	0.175
3	0.676	0.082	0.789	0.055	0.688	0.079
5	0.606	0.131	0.746	0.118	0.644	0.108
6	0.751	0.06	0.855	0.04	0.765	0.046
7	0.285	0.103	0.53	0.139	0.401	0.091
8	0.485	0.194	0.578	0.187	0.531	0.184
9	0.464	0.216	0.434	0.195	0.518	0.175

Table 67

Inter-model agreement descriptive statistics

Subject	Model	Kappa Score		CCC		MCC	
		Mean	Std	Mean	Std	Mean	Std
1	fcn	0.752	0.073	0.784	0.097	0.761	0.069
	resnet	0.629	0.097	0.712	0.076	0.636	0.093
2	fcn	0.398	0.128	0.534	0.17	0.426	0.138
	resnet	0.582	0.248	0.61	0.281	0.636	0.202
3	fcn	0.564	0.123	0.679	0.15	0.602	0.102
	resnet	0.465	0.129	0.523	0.156	0.509	0.102
5	fcn	0.786	0.054	0.897	0.043	0.796	0.047
	resnet	0.637	0.08	0.707	0.088	0.659	0.072
6	fcn	0.865	0.033	0.916	0.037	0.867	0.033
	resnet	0.825	0.037	0.894	0.025	0.829	0.036
7	fcn	0.583	0.13	0.823	0.092	0.631	0.096
	resnet	0.692	0.064	0.766	0.064	0.714	0.047
8	fcn	0.693	0.138	0.839	0.063	0.713	0.124
	resnet	0.313	0.14	0.362	0.132	0.390	0.113
9	fcn	0.442	0.225	0.575	0.23	0.514	0.174
	resnet	0.581	0.146	0.657	0.165	0.610	0.119

A2.6 Results statistics

A2.6.1 Measures

- DV = Performance for each maneuver and complexity level.
- IV = complexity level (1, 2, 3), MEMW, EMWP
- Manipulation check = perceived mental workload (Raw-TLX)

A2.6.2 Statistical tools

We employed the R programming language and the lme4 package (Bates et al., 2014) to conduct a linear mixed effects analysis. This approach was chosen because it accounts for the within-subject measurement of performance that violates the independence assumption of a linear regression model. We assessed the model's assumptions using the Performance package (Lüdecke et al., 2021).

A2.6.3 Linear mixed model procedure

Maximal random effect structure:

Specify a maximal LMEM based on the following criteria (Barr et al., 2013, p 275):

- “If a factor is between-unit, then a random intercept is usually sufficient.
- If a factor is within-unit and there are multiple observations per treatment level per unit, then you need a by-unit random slope for that factor.
- Exception: single observation for every treatment level of every unit, random slope variance would be completely confounded with trial-level error”

In our case, we have multiple observations of performance per subject and per complexity level. We assume that the effect of the manipulation, i.e., complexity might influence the performance and vary between subjects. Thus, we need to specify a random intercept and slope for subjects as influenced by the complexity of the maneuver.

Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using a Wald t-distribution approximation.

Table 68

Random effect structures

	(1|subject)	(1 | subject) + (1 | complexity)	(1 + complexity |subject)	(1 | subject) + (1 | complexity) + (1 | maneuver)
(Intercept)	2.869 (0.073)	2.850 (0.136)	2.879 (0.073)	2.856 (0.136)
	[2.724, 3.015]	[2.581, 3.120]	[2.733, 3.024]	[2.586, 3.126]
SD (Intercept subject)	0.133 (0.082)	0.158 (0.080)	0.519 (0.200)	0.168
	[0.039, 0.449]	[0.059, 0.424]	[0.244, 1.103]	
SD (Complexity subject)			0.268 (0.099)	
			[0.130, 0.551]	
Cor (Intercept~Complexity subject)			-0.958 (0.732)	
			[-1.000, 1.000]	
SD (Observations)	0.529 (0.042)	0.491 (0.039)	0.479 (0.040)	0.456
	[0.453, 0.617]	[0.419, 0.574]	[0.407, 0.565]	
SD (Intercept Complexity)		0.193 (0.102)		0.163
		[0.069, 0.545]		
SD (Intercept maneuver)				0.200
Num.Obs.	88	88	88	88
R2 Marg.	0.000	0.000	0.000	0.000
R2 Cond.	0.059	0.206	0.540	0.313
AIC	151.2	145.2	148.7	143.3
BIC	158.6	155.2	161.1	155.7
ICC	0.1	0.2	0.5	0.3
RMSE	0.52	0.47	0.45	0.43

Figure 42

Model check for the random structure (1 + Complexity |subject)

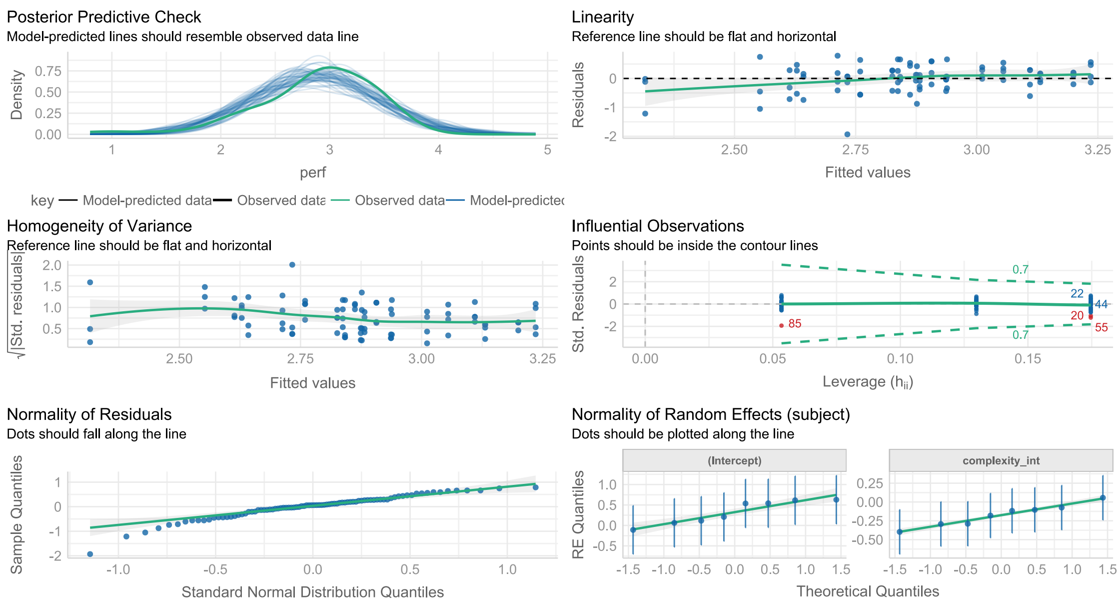


Table 69

Linear mixed model results

	ResNet			FCN	
	M1	M2	M3	M4	M5
Fixed Effects					
(Intercept)	3.358 p <0.001 (0.161)	3.434 p <0.001 (0.225)	3.424 p <0.001 (0.245)	3.304 p <0.001 (0.224)	3.313 p <0.001 (0.240)
	[3.037, 3.679]	[2.986, 3.882]	[2.937, 3.912]	[2.858, 3.750]	[2.835, 3.791]
Complexity	-0.256 p = 0.001 (0.075)	-0.257 p <p0.001 (0.074)	-0.258 p <0.001 (0.075)	-0.256 p = 0.001 (0.076)	-0.256 p = 0.001 (0.076)
	[-0.406, -0.106]	[-0.405, -0.109]	[-0.407, -0.108]	[-0.407, -0.104]	[-0.406, -0.105]
EMWP		-0.096 p = 0.642 (0.207)		0.074 p = 0.716 (0.203)	
		[-0.508, 0.315]		[-0.329, 0.477]	
MEMW			-0.078 p = 0.723 (0.221)		0.057 p = 0.796 (0.218)
			[-0.518, 0.361]		[-0.378, 0.491]
Random Effect					
SD (Intercept Pilot)	0.256 (0.222)	0.228 (0.238)	0.238 (0.232)	0.290 (0.218)	0.282 (0.220)
	[0.047, 1.395]	[0.029, 1.768]	[0.035, 1.604]	[0.066, 1.269]	[0.061, 1.297]
SD (complexity Pilot)	0.110 (0.113)	0.104 (0.118)	0.108 (0.115)	0.113 (0.112)	0.111 (0.113)
	[0.014, 0.831]	[0.011, 0.966]	[0.013, 0.871]	[0.016, 0.789]	[0.015, 0.820]
Cor (Intercept~Complexity subject)	-0.758 (1.774)	-0.631 (2.017)	-0.678 (1.886)	-0.787 (1.616)	-0.760 (1.662)
	[-1.000, 1.000]	[-1.000, 1.000]	[-1.000, 1.000]	[-1.000, 1.000]	[-1.000, 1.000]
SD (Observations)	0.479 (0.040)	0.481 (0.040)	0.481 (0.040)	0.480 (0.040)	0.480 (0.040)
	[0.407, 0.565]	[0.408, 0.566]	[0.408, 0.567]	[0.408, 0.566]	[0.408, 0.566]
Num.Obs.	88	88	88	88	88
Performance					
R2 Marg.	0.135	0.135	0.135	0.135	0.134

	ResNet		FCN	
	M1	M2	M3	M4	M5
R2 Cond.	0.252	0.266	0.263	0.264	0.265
AIC	146.3	149.4	149.4	149.5	149.4
BIC	161.1	166.8	166.7	166.9	166.8
ICC	0.1	0.2	0.1	0.1	0.2
RMSE	0.46	0.46	0.46	0.46	0.46

Note. M1 = performance ~ complexity, M2 = performance ~ EMWP(ResNet) + complexity, M3 = performance ~ MEMW(ResNet) + complexity, M4 = performance ~ EMWP(FCN) + complexity, M5 = performance ~ MEMW(FCN) + complexity. Random structure = (1 + Complexity |subject). Values represent: estimate, p.value (std.error) [conf.low, conf.high].

A2.6.4 Pearson's correlation analyses

Table 70

Pearson's correlation analyses

Perceived Mental Workload	Estimated Mental Workload	r	95% CI	t(86)	p
RAW-TLX	ResNet - EMWP	0.24	[.03, .42]	2.25	.027*
RAW-TLX	ResNet - MEMW	0.22	[.01, .41]	2.07	.041*
RAW-TLX	FCN - EMWP	0.25	[.04, .44]	2.39	.019*
RAW-TLX	FCN - MEMW	0.25	[.04, .43]	2.37	.020*

A3 Chapter 4

A3.1 Experimental setup – study 1

Figure 43

Study 1 - room configuration

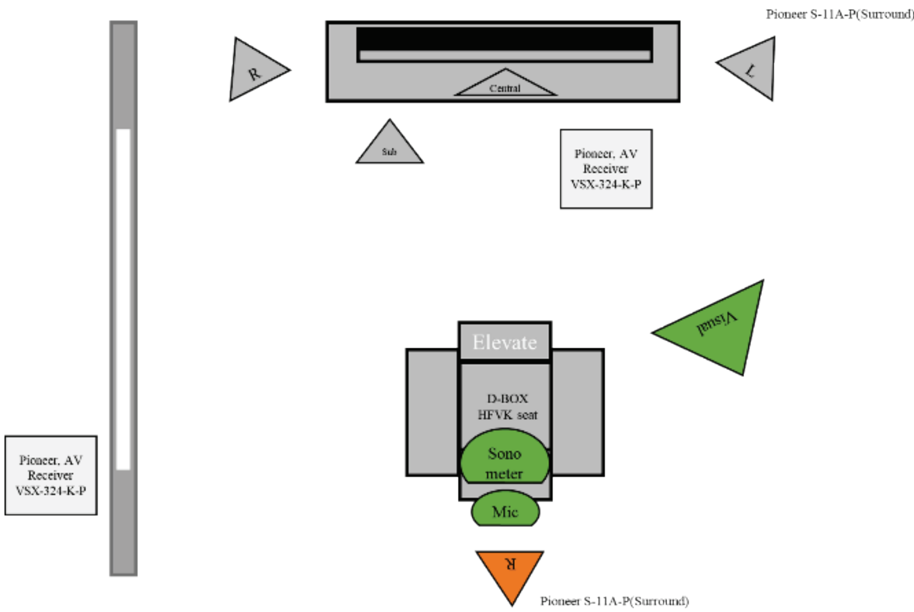
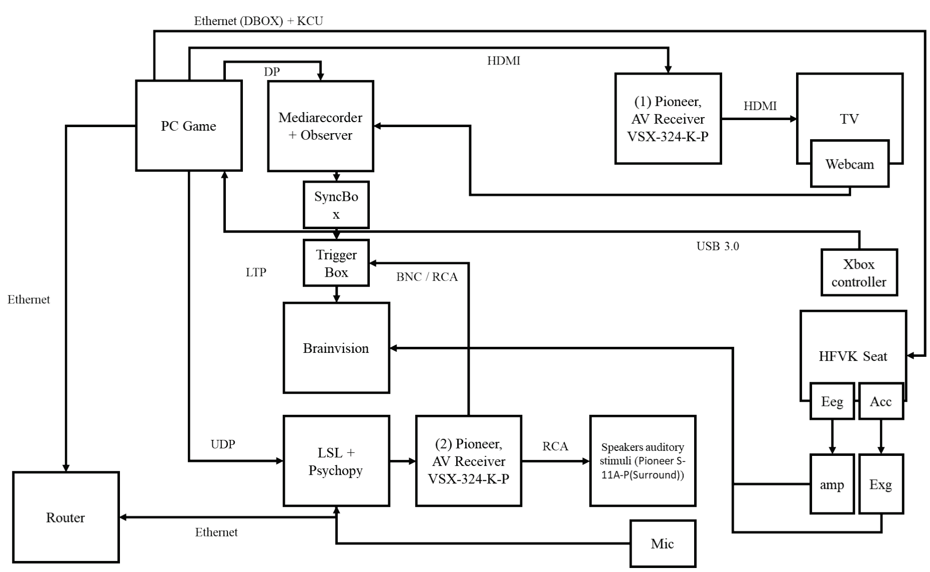


Figure 44

Study 1 - data collection infrastructure and synchronisation



A3.2 Experimental setup – study 2

Figure 45

Study 2 - data collection infrastructure and synchronisation