**(055)**

# Enhancing Operator Engagement during AI-assisted Manufacturing Work Using Optimal State Deviation Feedback System

L. Couture [1], M. Passalacqua [2], L. Joblot [3], F. Magnani [4], R. Pellerin [2] and P. M. Léger [1]
[1] HEC Montréal, Tech3Lab, 3000 Chemin de la Côte-Sainte-Catherine, H3T 2A7, Montréal, Canada
[2] Polytechnique Montréal, 2500 Chemin de Polytechnique, H3T 1J4, Montréal, Canada
[3] Arts et Métiers Institute of Technology, LISPEN / HESAM University / UBFC, Chalon-Sur-Saône, France
[4] Université Aix-Marseille, CERGAM, Centrale Méditerranée, 38, rue Frédéric Joliot-Curie, 13013, Marseille, France
Tel.: + 15149981557
E-mail: loic.couture@hec.ca

**Summary:** The integration of Artificial Intelligence (AI) in manufacturing is shifting the focus of operators from manual labor to cognitive supervision roles. While this transition demands more engagement from operators, the less stimulating nature of monitoring tasks has, paradoxically, reduced operator involvement, consequently presenting new challenges in performance maintenance. Addressing this issue, our research adopted an iterative design science methodology to create a biocybernetic system that aims to enhance operator engagement in their evolving workplace. This system leverages physiological signals to intuitively display how much an operator's engagement level deviates from an ideal state, ensuring operators stay aware of their psychophysiological state of engagement and can quickly adjust to any decreases in engagement. In this paper, we detail the 4-step process that led to the development of the first version of the system. Capitalizing on the physiological differences observed in manufacturing operators during "high" and "low" engagement scenarios, we defined a task-specific Optimal State Deviation Index (OSDI) formula. This formula enabled us to predict participants' engagement states with an 80.95 % success rate in our testing dataset.

**Keywords:** Biocybernetic system, Manufacturing, Engagement, Automation, Design science, Artificial intelligence.

## 1. Introduction

AI-driven automation is transforming manufacturing operators' roles, shifting their work from manual work to supervising systems [1], which can lead to less stimulating tasks, adversely impacting operator engagement and performance [2]. Given the prevalent risk of occupational injuries associated with manufacturing work [3], it appears essential for operators to maintain an optimal engagement state. Specifically, operators must avoid excessive vigilance, which can increase fatigue over time or lead to cognitive tunneling, a state in which operators adopt a narrow focus and neglect other important information [4]. Operators must also avoid cognitive underload, which can result in mind wandering and inattention [5]. In addressing the issue of maintaining an optimal engagement state, very little research has explored how new technologies can effectively improve operator engagement in manufacturing.

However, the work of Demazure et al. [6] is particularly promising in this regard. Their research demonstrated the potential of using real-time engagement level feedback to significantly improve users' attentiveness. Our study seeks to adapt this approach for manufacturing, aiming to develop a tool to help operators maintain optimal engagement levels.

The structure of this paper is outlined as follows. Section 2 delves into the reasons for developing a new biocybernetic system tailored for manufacturing. Section 3 is dedicated to detailing the iterative design science methodology that was employed to create the system. The results that influenced the system's design are detailed in Section 4. Finally, we present our concluding remarks, along with a discussion of the current system's limitations in Section 5.

## 2. Background

The integration of automated systems offers significant advantages for industrial applications. Nonetheless, it is important to acknowledge that most of these automated systems have yet to achieve perfection in terms of system reliability [7]. Consequently, in instances where a system's reliability is not absolute, it is prudent for enterprises to deploy human operators. These operators play a crucial role in monitoring automated systems' functionality, enabling the early detection of anomalies and facilitating timely intervention to rectify such occurrences. However, monitoring automated systems can present several human challenges, including a decrease in vigilance over time [8] and low monitoring performance [2]. This decline is attributed to both cognitive overload, which can result in cognitive fatigue and cognitive tunneling [9], and cognitive underload, which can cause mind wandering, low motivation, and increased distraction [10]. Therefore, one way to tackle this issue is to ensure the operator can balance their level of engagement throughout the monitoring task.

In this context, the work of Demazure et al. [6] seems particularly promising as it offers a passive system that informs the operator of their level of engagement in real-time. This feature not only keeps operators aware of their mental state in real-time but also enables them to make immediate adjustments as needed. The solution proposed by Demazure utilizes electroencephalography (EEG) signals to provide a real-time, intuitive display of the operator's engagement level through a color gradient display. Karran et al. [11] employed Demazure et al.'s system and demonstrated that continuously showing engagement levels to operators notably enhanced sustained attention during long monitoring tasks. This was evidenced by increased EEG wave coherence recorded for participants who received continuous engagement feedback. In contrast, participants who did not receive engagement feedback and those who only received engagement feedback after critical disengagement thresholds were reached reported low EEG wave coherence. While these results appear encouraging, a notable challenge with this solution is the necessity of an accurate measurement of engagement, which can be particularly difficult in manufacturing settings.

Numerous physiological tools have been used in the literature to measure task engagement, including eye-tracking [12], electroencephalography (EEG) [6], electrodermal activity (EDA) [13], and heart rate variability measures (HRV) [14]. Although eye-tracking and EEG methods are well-established in the literature for assessing engagement, their practical application in manufacturing faces significant challenges. The primary issue with these techniques is their limited adaptability to the dynamic nature of manufacturing environments. Operators in such settings are frequently mobile and engage with their surroundings in a 360-degree manner. This constant movement and the need to interact with a wide-ranging environment render both eye-tracking and EEG methodologies less feasible due to their inherent requirement for relative stability and controlled observation conditions. EDA is typically measured on the palm of the hand, which could constrain operators in their work. However, HRV can accurately be assessed during operator movement, making it a potential choice for a manufacturing setting [15]. HRV is defined as the variation of time intervals between consecutive heartbeats [16] and is mainly used as a measure of the activation of the autonomous nervous system [17]. There is, however, some debate regarding the interpretation of HRV measures [17, 18], which raises questions regarding the viability of using this metric to assess task engagement.

This ambiguity makes Moray and Inagaki's approach [19] particularly appealing. Their method evaluates monitoring performance by contrasting actual operator performance to an optimal standard. From this perspective, for any specific task, it seems feasible to establish a performance metric by initially recording the responses of an operator in a high-performance scenario and comparing it to a low-performance scenario. Therefore, when we want to assess operator engagement, a potential approach would be to establish an engagement metric by comparing physiological responses recorded in highly engaging scenarios with those from a minimally engaging scenario, using contrast to construct a reliable measure of engagement for this particular task. Additionally, since increasing the level of automation has been shown to be the source of lower engagement [20], it seems possible to use the levels of automation to induce different levels of engagement in a manufacturing context.

Hence, to maintain optimal engagement levels of manufacturing operators within their dynamic work environments, our proposal involves developing a new biocybernetic system inspired by the research of Demazure et al. [6] but tailored to the manufacturing context. Rather than depending on exact engagement metrics and measurements, our system follows a methodology similar to Moray and Inagaki [19], leveraging physiological indicators that differentiate between optimal and suboptimal engagement states. A significant advantage of this approach is its adaptability to complex settings like manufacturing, where constraints exist concerning the feasibility of certain physiological measurements, such as eye-tracking and EEG.

## 3. Methods

We used a design science methodology to develop an optimal state deviation feedback system involving a four-step process that included three studies (see Table 1). The first two steps were dedicated to identifying physiological markers that could characterize the reduction of operators' engagement during a specific task and developing a biocybernetic system. The last two steps were dedicated to evaluating different features of the biocybernetic system, i.e., the display modality and the scaling method.

### 3.1. Step 1 – Collect Data

In the first step, we collected physiological and perceptual data from participants in more and less engaging manufacturing situations. We recruited 22 students (age = 21.62±3.17; men = 14) for a within-subject experiment, in which they twice performed a quality control and assembly task on a simulated assembly line[4]. All participants provided a signed consent in-line with the University ethics committee (project # 2023-5058) and were

[4] For an overview of the experimental setup: https://youtu.be/xtcpxqcyz8k

compensated with the sum of 40 euros. The task, explained in more detail in [2], required participants to detect errors on partially assembled snowshoes and complete the assembly by fixing the binding to the base at its pivot point (see Fig. 1). In the "less engaging" condition, we automated the participants' decision-making, equipping them with a fully reliable error detection system that indicated to the operator whether or not a snowshoe had a defect. In the "more engaging" condition, participants had to manually detect errors before assembling the snowshoes. During each task, a total of 30 snowshoes had to be assembled by the participants, with six being defective. Participants realized the task once with automated support and once without automated support, with condition order being randomly assigned and counterbalanced. During the task, we collected physiological data using a Hexoskin vest [21], recording heart rate, respiratory rate, and acceleration data. We also collected perceived cognitive absorption, vigor, and dedication using the Utrecht Work Engagement Scale (UWES) [22], which was collected post-task. The raw physiological data from the Hexoskin was pre-processed and synchronized using the COBALT Photobooth software [23]. The list of physiological and self-reported data collected can be found in Table 2.

**Table 1.** Methodology employed to design the biocybernetic system.

| Step | Step 1 | Step 2 | Step 3 | Step 4 |
|---|---|---|---|---|
| **Title** | *Collect data* | *Identify markers* | *Display validation* | *Scaling validation* |
| **Description** | Study 1: Collection of Physiological Data in Scenarios with Varied Engagement Levels | Identify physiological markers of engagement & design the system | *Study 2:* Validating multiple display modalities of engagement | *Study 3:* Validating multiple index scaling methods |
| **Experimental design** | Between-subject | – | Within-subject | Between subject |
| **Conditions** | No automation<br>Automation | – | Discrete color gradient (3 shades of color)<br>Continuous color gradient (100 shades between green and red) | Min/Max since the beginning of the task<br>Min/Max of training data<br>Min = $25^{th}$ & Max = $75^{th}$ since the beginning of the task |
| **Experimental manipulation** | Manufacturing Q&A and assembly tasks using snowshoes. | Feature extraction using a logistic regression model<br>Validation with LOOCV | Fully automated manufacturing Q&A and assembly tasks using images of snowshoes. | Fully automated manufacturing Q&A and assembly tasks using images of snowshoes |
| **Data** | Collected physiological data (bpm, breath rate, motion) and perceived work engagement (UWES) | Task 1 & Task 2 data from step 1 | 10 minutes semi-directed interviews | Five questions questionnaire |
| **Participants** | 22 participants | - | 3 participants | 10 participants |

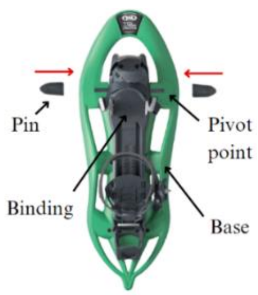


**Fig. 1**. Product used in the manufacturing task.

### 3.2. Step 2 – Identify Markers & System Design

In the second step, we began by validating our primary assumption that the condition with automation was less engaging than the manual condition. Due to a noticeable learning effect between the first and second tasks, primarily manifested in performance improvements, we chose to focus exclusively on the results obtained from the first task, where no learning effects could affect perception. We compared the perceived absorption, dedication, and vigor scores between automated and manual conditions using the Mann-Whitney-Wilcoxon Test, which is suitable for comparing non-parametric independent samples.

To categorize the data, we assigned labels of "high" or "low" engagement to arrays of data, depending on the condition experienced by the participant. Data originating from the automated task was labeled as "low engagement," while data from the manual task was labeled as "high engagement". We then defined a task-specific optimal state deviation index (OSDI) using the three physiological variables with the highest estimated weights in the logistic regression model used to predict the probability of a participant being more engaged in the task. The whole dataset (Task 1 & Task 2) was used to develop the formula. The formula represents a weighted sum, where each coefficient corresponds to the respective variable's estimated power to predict if a participant is in a "high" or "low" state of engagement. The formula is based on 30-second data windows.

$$OSDI = (435.7\ \text{motion}_{\text{std}}) - (175.4\ \text{motion}_{\text{mean}}) + (0.78\ \text{breathingRate}_{\text{std}}) \qquad (1)$$

**Table 2.** List of collected variables

| Type of data | Measure | Description |
|---|---|---|
| Physiological data | Beats per minute | Number of beats per minute |
| | SDNN | Standard deviation of NN intervals |
| | LF | Power of the Low-frequency band (0.04-0.15 Hz) ($ms^2$) |
| | HF | Power of the High-frequency band (0.15-0.4 Hz) ($ms^2$) |
| | LF/HF | Ratio of Low-to-High frequency power |
| | Breathing Rate | Number of respirations per minute |
| | Minute Ventilation | Respiratory volume per minute (L/min) |
| | Cadence | Number of steps per minute |
| | Motion | Norm of the 3D acceleration vector (G) |
| Self-reported measures | Absorption score | Perceived absorption |
| | Vigor score | Perceived vigor |
| | Dedication score | Perceived dedication |

To validate the formula, we employed the leave-one-out cross-validation (LOOCV) using the OSDI in a logistic model to predict if a participant's engagement during a task was "higher" or "lower". The same dataset was used for this validation step. We then developed a biocybernetic system on Python that employs the OSDI formula to calculate the index in real-time, scale it, and visually represent it as a color gradient (see Fig. 2). The system received pre-processed physiological data every second (1 Hz) from the Hexoskin vest. It calculated the engagement index using the OSDI formula based on the last 30 seconds' data. The first prototype (and the one used for the next step) scaled the OSDI between [0-100] using the minimum and the maximum values since the beginning of the task.

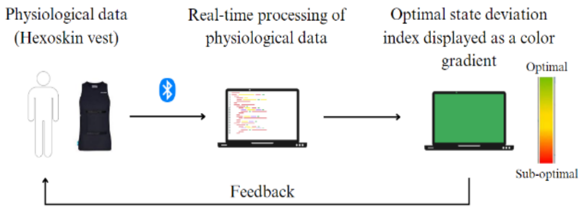


**Fig 2.** Overview of the biocybernetic system

### 3.3. Step 3 – Display Validation

In the third step, we assessed whether representing the index through a continuous color gradient (100 shades) or a discrete color gradient (3 colors) was more effective in conveying participants' engagement levels. We recruited three participants for a within-subjects pilot test. Each participant completed a low-fidelity version of the automated assembly task twice (using printed images of snowshoes instead of real snowshoes), experiencing the feedback system in both formats. After completing each task, participants underwent a 5-minute semi-directed interview. During this interview, they were asked about their perceptions of the system's impact on their engagement, the potential distractions caused by the system, and its effectiveness in representing their engagement levels. Positive and negative statements in each category were compiled and analyzed, making the decision to retain the continuous color gradient.

### 3.4. Step 4 – Scaling Validation

In the fourth step, we aimed to identify the most effective method for scaling the index. We tested three scaling methods: (i) dynamically adjusting the minimum and maximum values based on the minimum and maximum values recorded since the beginning of the task, (ii) using the minimum and maximum values of the training dataset, measured with formula (2) to exclude outliers, and (iii) dynamically setting the minimum and maximum values respectively to the 25th and 75th percentile of the data since the beginning of the task.

$$MIN/MAX = OSDI_{mean} \pm 3 * OSDI_{std} \qquad (2)$$

We performed a between-subjects experiment with 10 participants who each completed the same low-fidelity version of the manufacturing task while being assisted by the system in one of its three possible formats (using printed images of snowshoes instead of real snowshoes). After completing the task, participants were asked to rate the representativeness, interpretability, and distractive nature of the color display on a scale from 0 to 100.

## 4. Results

The one-sided Mann-Whitney-Wilcoxon Test used **for step two** revealed a statistically significant difference in perceived absorption scores between manual and automated conditions (p =.03; d = 0.83), suggesting that the reported absorption scores tend to be lower in the automated condition compared to manual condition. This result supports our primary assumption that the automated condition was less engaging than the manual condition. No significant differences were found between conditions for dedication (p =.40; d = -.82) and vigor (p =.82; d = -.43) subscales of UWES during task 1 (see Fig. 3).

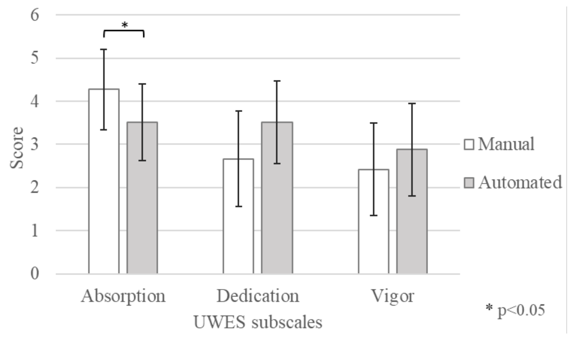


**Fig. 3.** Task 1 UWES Questionnaire Results: Participant Response Analysis.

Using the OSDI formula to predict if a participant was in a "high" or "low" state of engagement in a logistic regression model, we achieved 81.31 % accuracy on the training set and 80.95 % on the testing set, as confirmed through leave-one-out cross-validation. **For step three**, where we assessed the display modality, we employed a qualitative labeling technique to categorize interview statements into three themes: effect on perceived engagement, distraction, and representativeness. The number of statements in each category was then compiled (see Table 3), showing that the discrete color gradient was more distracting (0 positive, six negative statements) than the continuous color gradient (2 positive, 0 negative statements).

**Table 3.** Compilation of qualitative statements on continuous and discrete color gradients.

| | **Perceived effect on engagement** | | **Distraction** | | **Representa-tiveness** | |
|---|---|---|---|---|---|---|
| | (+) | (-) | (+) | (-) | (+) | (-) |
| Continuous | 5 | 0 | 2 | 0 | 2 | 2 |
| Discrete | 2 | 1 | 0 | 6 | 0 | 3 |

**In step four**, the self-reported data from questionnaires revealed that all methods were equally easy to interpret and not distracting. However, the scaling method (ii) utilizing the minimum and maximum values from the training dataset proved to be more representative, with a mean score of $93.33\ \% \pm 6.24\ \%$. This was in contrast to the scaling method (i), which was based on the minimum and maximum values since the beginning of the task ($mean\ =\ 57.33 \pm 12.28\ \%$), and method (iii) which was based on percentiles ($mean\ =\ 45.5 \pm 14.5\ \%$), as illustrated in Fig. 4. Based on these analyses, we concluded that the continuous color gradient and scaling method, which utilized the minimum and maximum values of the training dataset, i.e., method (ii), are preferred options for any future work.

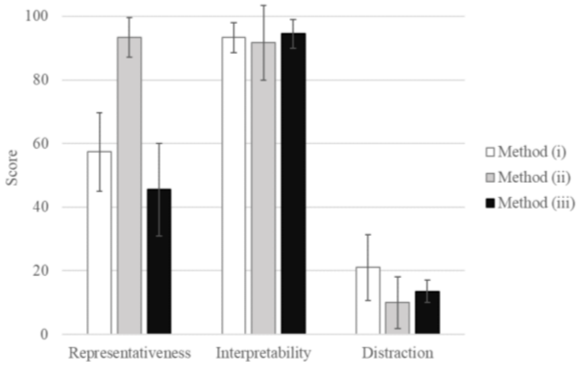


**Fig. 4.** Scaling method comparison: Evaluating Representativeness, Interpretability, and Distraction through Questionnaire Scores.

## 5. Conclusions

This study employed a design science methodology to create an optimal state deviation feedback system designed to help manufacturing operators stay engaged in their workplace. The task-specific optimal state index was developed using physiological data collected during a simulated manufacturing assembly task, achieving 80.95 % accuracy in predicting the engagement state of the testing set. We assessed two display modalities and three scaling methods to inform our design. The final design utilized a continuous color gradient calibrated based on the lowest and highest values of the training set. A subsequent study was conducted to test this advancement in a broader scale, which will be discussed in forthcoming scientific publications.

It is essential to acknowledge certain limitations inherent in this system. First, our assessment of engagement relied solely on self-reported data. Ideally, employing real-time physiological monitoring tools, like EEG, would have enhanced the validation of the measured engagement levels but would have been much more intrusive than the Hexoskin vest we used. Additionally, it should be noted that while the leave-out samples were not employed in training the predictive models, they were utilized in creating the OSDI formula. As a result, the model's effectiveness for new participants might not be as robust as measured in this study. Finally, it is important to note that the formula used in this system strongly depends on the task and is specifically tailored to the context of our study. This means that the OSDI formula may not yield reliable results in different contexts and, therefore, should not be applied to other scenarios without appropriate modifications and validation.

## References


[1]. R. Parasuraman, C. D. Wickens, Humans: still vital after all these years of automation, in Decision Making in Aviation, *Routledge*, 2017, pp. 251-260.

[2]. M. Passalacqua, R. Pellerin, E. Yahia, F. Magnani, F. Rosin, L. Joblot, P. M. Léger, Practice with less AI makes perfect: partially automated AI during training leads to better worker motivation, engagement, and skill acquisition, *International Journal of Human-Computer Interaction*, 2024 (accepted).

[3]. Occupation Groups with the Highest Incidence Rate of Nonfatal Occupational Injuries and Illnesses per 10,000 Full-Time Workers in the U.S. in 2020, *Bureau of Labor Statistics*, 2021.

[4]. S. Pooladvand, S. Hasanzadeh, Impacts of stress on workers' risk-taking behaviors: cognitive tunneling and impaired selective attention, *Journal of Construction Engineering and Management*, Vol. 149, Issue 8, 2023, 04023060.

[5]. F. Dehais, A. Lafont, R. Roy, S. Fairclough, A neuroergonomics approach to mental workload, engagement and human performance, *Frontiers in Neuroscience*, Vol. 14, 2020.

[6]. T. Demazure, et al., Enhancing sustained attention, *Business & Information Systems Engineering*, Vol. 63, Issue 6, 2021, pp. 653-668.

[7]. K. Wang, J. Lu, S. Ruan, Y. Qi, Continuous error timing in automation: the peak-end effect on human-automation trust, *International Journal of Human-Computer Interaction*, 2022, pp. 1-13.

[8]. M. Körber, A. Cingel, M. Zimmermann, K. Bengler, Vigilance decrement and passive fatigue caused by monotony in automated driving, *Procedia Manufacturing*, Vol. 3, 2015, pp. 2403-2409.

[9]. P. A. Desmond, P. A. Hancock, Active and passive fatigue states, in Stress, Workload, and Fatigue, *CRC Press*, 2000, pp. 455-465.

[10]. S. Conte, D. Harris, J. Blundell, Evaluating the impact of passive fatigue on pilots using performance and subjective states measures, in *Proceedings of the International Conference on Human-Computer Interaction*, 2023, pp. 21-36.

[11]. A. J. Karran, et al., Toward a hybrid passive BCI for the modulation of sustained attention using EEG and fNIRS, *Frontiers in Human Neuroscience*, Vol. 13, 2019-November-06.

[12]. R. N. Roy, A. Bovo, T. Gateau, F. Dehais, C. P. C. Chanel, Operator engagement during prolonged simulated UAV operation, *IFAC-PapersOnLine*, Vol. 49, Issue 32, 2016, pp. 171-176.

[13]. U. Kale, J. Rohács, D. Rohács, Operators' Load monitoring and management, *Sensors*, Vol. 20, Issue 17, 2020, 4665.

[14]. S. Pütz, A. Mertens, L. Chuang, V. Nitsch, Physiological measures of operators' mental state in supervisory process control tasks: a scoping review, *Ergonomics*, 2023, pp. 1-30.

[15]. S. Lee, H. Kim, D.-H. Kim, M. Yum, M. Son, Heart rate variability in male shift workers in automobile manufacturing factories in South Korea, *International Archives of Occupational and Environmental Health*, Vol. 88, Issue 7, 2015, pp. 895-902.

[16]. R. Mccraty, F. Shaffer, Heart Rate Variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk, *Global Advances in Health and Medicine*, Vol. 4, Issue 1, 2015, pp. 46-61.

[17]. F. Shaffer, R. McCraty, C. L. Zerr, A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability, *Frontiers in Psychology*, Vol. 5, 2014, 1040.

[18]. G. E. Billman, The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance, *Front. Physiol.*, Vol. 4, 2013, 26.

[19]. N. Moray, T. Inagaki, Attention and complacency, *Theoretical Issues in Ergonomics Science*, Vol. 1, Issue 4, 2000, pp. 354-365.

[20]. D. Manzey, J. Reichenbach, L. Onnasch, Human performance consequences of automated decision aids: the impact of degree of automation and system experience, *Journal of Cognitive Engineering and Decision Making*, Vol. 6, Issue 1, 2012, pp. 57-87.

[21]. Hexoskin, https://www.hexoskin.com

[22]. W. B. Schaufeli, A. B. Bakker, M. Salanova, Utrecht Work Engagement Scale, Educational and Psychological Measurement, *Utrecht University*, 2003.

[23]. P.-M. Léger et al., Caption and observation based on the algorithm for triangulation (COBALT): Preliminary results from a beta trial, in Information Systems and Neuroscience, NeuroIS 2022, *Springer*, 2022, pp. 229-235.