*Artificial Intelligence and Social Computing,* Vol. 122, 2024, 41–51
https://doi.org/10.54941/ahfe1004640

# Evaluation of a Scale to Assess Subjective Information Processing Awareness of Humans in Interaction With Automation & Artificial Intelligence

**Tim Schrills, Marvin Sieger, Marthe Gruner, and Thomas Franke**

University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany

## ABSTRACT

Subjective Information Processing Awareness (SIPA) describes how users experience the extent to which a system enables them to perceive, understand and predict its information processing. With the rising interdependence of information processing in Human-AI interaction, research methods for assessing user experience in automated information processing are needed. The objective of the present research was the empirical evaluation of the SIPA scale as an economical method to assess SIPA, as well as the construction of and comparison with a version in plain language. To this end, two empirical studies were conducted ($N_{S1} = 317$, $N_{S2} = 230$) to enable scale analysis. Results showed that the SIPA scale achieves excellent reliability and expected correlations with connected constructs, e.g., trust and perceived usefulness of AI systems. In addition, no benefits of a plain language variant were found. Based on the results, the SIPA scale appears to be a promising tool for examining user experience of systems with automated information processing.

**Keywords:** Explainability, Trust, Human-centered AI, Human-AI cooperation, User experience measures

## INTRODUCTION

The growing incorporation of artificial intelligence (AI) systems in everyday life (Zhang et al., 2021), work (Jarrahi, 2018), and key sectors such as medicine (Rajpurkar et al., 2022) or transport (Fatemidokht et al., 2021) changes how human users process information and how they interact with automated systems. One major challenge for integrated information processing in Human-AI teams is linked to a known problem in human-automation interaction: loss of Situation Awareness (SA), which can negatively impact humans' ability to work with and monitor automated systems (Endsley, 1995). Users might not be able to access necessary information relevant to determine the right course of action, e.g., making a diagnosis, therefore inhibiting users' ability to control the system. That is, to maintain the ability to act adequately in Human-AI interaction, awareness of how information is processed by intelligent systems appears to be essential for users and impacts user experience in automation. As a result, recent research emphasizes the importance of transparent systems and, in the context of explainable AI (XAI, see Arrieta et al., 2019), try to develop solutions for transparency. Previous research demonstrated that explanations might affect user experience, which is crucial for users when choosing and controlling their actions (i.e., action regulation) and for assessing their own abilities.

However, most research in the XAI literature focuses on the direct evaluation of explanations (c.f. Explanation Satisfaction Scale, see Hoffman et al., 2019). In contrast, assessing users' experience of a system's ability to enable their information processing awareness (i.e., related to the concept of SA) appears to allow the research of underlying cognitive mechanisms. While corresponding methods exist (see Subjective Information Processing Awareness Scale (SIPA); Schrills et al., 2022), a psychometric evaluation of scale quality criteria is required to substantiate broader use in Human-AI interaction research. In addition, to avoid a methodological limitation, the applicability of the scale for a broad diversity of users needs to be evaluated.

The objective of the present research was to elicit psychometric data of the SIPA scale for assessing user experience in Human-AI interaction and examine the scale's appropriateness for diverse user groups. To this end, we examined ways to further improve the accessibility of the scale, that is, formulating the scale items with more simplified wording (i.e., plain language). Therefore, data from two studies in which participants interacted with intelligent systems – with information systems in Study 1(S1) contact tracing and Study 2 (S2) music recommendation – was analyzed.

## BACKGROUND

Situation awareness has been established as a key construct for understanding interactions with automated systems (e.g., Popken & Krems, 2011 & Endsley, 2018). Endsley (1995) defines SA as the ability of humans to (1) perceive relevant elements in the environment, (2) understand their relationship, and (3) predict the development of the situation. On the one hand, low SA can limit users' ability to intervene in automated processes when problems arise. On the other hand, only with a sufficiently accurate situational model (Endsley, 2015), users may recognize constellations in which an automated system may lack critical information or pursue different goals. The importance for designers to make information processing sufficiently transparent is supported by Situation Awareness-Based Transparency (SAT, Chen et al., 2018). This research identifies what systems can do to maintain people's automation-related SA. For example, conveying current system goals or describing potential limits are suggested methods to improve Level 1 of SA (i.e., perception). In SAT, the requirements are referred to as a "subset" of SA.

While existing research proposes technical solutions to enhance SA in automation, e.g., SAFE-AI already focuses on the human users and SA-based needs in terms of communicated information (Sanneman & Shah, 2022). Accordingly, existing research may address SA not entirely, but an automation-specific "subset". This subset refers to the automatic processing of information and the associated awareness. Looking at AI as a form of automated information processing, we refer to this subset of information processing as information processing awareness (IPA). IPA is the cognitive state in which users can perceive, understand, and predict the processing of information of an automatic system.

Information processing awareness, parallel to SA, describes, in three levels, how users achieve the ability to effectively interact with information processing systems, namely by perceiving, understanding, and projecting the situation's development. For example, decision support systems are designed to support a user's ability to build sufficient information processing awareness but may hamper it if not designed properly. That is, system properties determine users' ability to control (or cooperate with) the system. These properties are directly connected to the three levels of SA and are the system's transparency, understandability, and predictability (as defined also in Schrills et al., 2022), which also represents three facets of SIPA.

Low levels of SA in automated systems may cause an "out-of-the-loop unfamiliarity", which can also be associated with low levels of trust (Lorenz et al., 2001). Trust as a central variable in Human-AI interaction is discussed and analyzed in existing frameworks, e.g., by Chiou & Lee, 2021, allowing us to link this construct to IPA. For example, previous research highlights that trust in AI systems is strongly connected to the goals users attribute to a system (Ferrario et al., 2020). When users experience high alignment between their own and the system's intentions, trust levels may rise (Lyons et al., 2023). Since awareness of goals is integral to IPA, experiencing high levels of SIPA might be essential to also experience high levels of trust. Additionally, for cooperation between users and intelligent systems, predictability and directability are central criteria (Chiou & Lee, 2021; Klein et al., 2004). Parallel to SA, a valid IPA is needed for users to be able to select actions that allow control over the system, i.e., users' ability to correctly predict how a change in information will change the output of the system's information processing is important for useful collaborative systems (Klein et al., 2004). Accordingly, when users report low levels of SIPA, their evaluation of the system's usefulness should also decrease. Users' confidence in choosing the correct actions or using the system may also be affected by low SIPA levels.

## THE SIPA SCALE AND ITS ADOPTION IN PLAIN LANGUAGE

Based on the three levels of SIPA (i.e., transparency, understandability, and predictability), the corresponding scale consists of three subfacets. Each facet is assessed with two items (see Table 1). Responses are captured on a Likert scale from 1 (completely disagree) to 6 (completely agree). For the data analysis, the mean score of all six items can be calculated to obtain the overall individual SIPA score. Optionally, mean values for transparency (items 01 and 02), understandability (items 03 and 04), and predictability (05 and 06) can be calculated separately (see Schrills et al., 2022). Importantly, the wording of single items is not to be modified, but rather the instruction may be adapted according to the system being evaluated.

**Table 1.** English items of the SIPA scale and instructions (as published in Schrills & Franke, 2023), answered from 1 = completely disagree to 6 = completely agree.

| | |
|---|---|
| The following questionnaire deals with your experience in the interaction with the system. Information refers to all data that the system can work with. Result refers to the output of the system, which is presented at the end of the system's information processing. | |
| 01 | I am aware of what information the system can collect. |
| 02 | I have access to the same information as the system. |
| 03 | I understand how the system calculates the result based on the information. |
| 04 | I understand what the system does with the information. |
| 05 | I am aware of how the system reacts to a change in information. |
| 06 | I have a clear idea which information the system would use to come to a different result. |

The German version of the scale was constructed along with the English version (both scales were developed by the authors at the same time). After that, the scale and its translations were independently assessed by two experts with professional background in translation. The German items can be found in Table 2.

**Table 2.** German items of the SIPA scale and instructions (as published in Schrills & Franke, 2023), answered from 1 = stimmt gar nicht 6 = stimmt völlig.

| | |
|---|---|
| Der folgende Fragebogen bezieht sich auf Ihr Erleben in der Interaktion mit dem System. Mit Informationen sind dabei alle Daten gemeint, mit denen das System arbeiten kann. Mit Ergebnis ist die Ausgabe des Systems gemeint, welche am Ende der Informationsverarbeitung des Systems dargestellt wird. | |
| 01 | Es war für mich transparent, welche Informationen durch das System gesammelt wurden. |
| 02 | Die Informationen, die das System erfassen konnte, waren für mich erkennbar. |
| 03 | Es war verständlich für mich, wie die gesammelten Informationen zum Ergebnis geführt haben. |
| 04 | Die Informationsverarbeitung des Systems war für mich nachvollziehbar. |
| 05 | Mit den mir zugänglichen Informationen war das Ergebnis vorhersehbar für mich. |
| 06 | Die Informationsverarbeitung des Systems war vorhersagbar für mich. |

However, since prior research endorsed having three subfacets (Schrills et al., 2022), no subfacet was dropped. However, the high complexity of the wording could cause problems when differentiating items, especially regarding understandability and predictability. We identified criteria that guided the development of a plain language variant. These included, among others: 1. The use of short sentences, 2. The use of short words from everyday language, 3. The division of complex thoughts into just one per sentence, 4. The assumption that no prior knowledge can be assumed, 5. The use of active verbs. Based on those principles, a first version of the simplified scale was developed during an explorative workshop. The aim of this first version was to serve as a basis for discussion for the development of the final items. A second iteration of the SIPA-P was then developed collaboratively in a larger expert workshop. To this end, the basic concept of simple language was first explained to the participants before they were asked to revise individual items in small groups and then move on to the next item. The results of this phase were then presented and discussed in plenary to finalize the final version. The revised items of the SIPA-P scale can be found in Table 3.

**Table 3.** Items of the SIPA-P scale (SIPA in plain language).

| | |
|---|---|
| 01 | I am aware of what information the system can collect. |
| 02 | I have access to the same information as the system. |
| 03 | I understand how the system calculates the result based on the information. |
| 04 | I understand what the system does with the information. |
| 05 | I am aware of how the system reacts to a change in information. |
| 06 | I have a clear idea which information the system would use to come to a different result. |

## EMPIRICAL EVALUATION

Based on previous research on the SIPA scale, the objective of our empirical research was to 1) examine the psychometric qualities of the SIPA scale and 2) evaluate the SIPA-P scale in comparison to the SIPA scale.

To this end, we conducted two online studies. For the first study (S1), the Corona Warn App (CWA; Corona-Warn-App, 2023) was used as an example of an automatic information processing system for which SIPA can be measured. The CWA is a medical application for tracing COVID-19 contacts that was widely used in Germany during the pandemic. Its main feature is the ability to automatically send alerts based on contact status changes to warn users about potential virus infections. The second study (S2) that was conducted focused on the "create station" feature of the music streaming service Spotify.

For S1, a sample of $N = 317$ CWA users was recruited using student e-mail distribution lists, social media platforms, and internal university forums. The study was conducted in German. Demographic data such as age, gender, and level of education were not collected as they were irrelevant for any of the evaluations and in the context of health behavior, the highest level of anonymization was intended. After completing the questionnaire, psychology, and media informatics students were able to receive course credit as compensation for participation. Other participants did not receive any compensation. The study was preregistered under: https://doi.org/10.17605/OSF.IO/X5YD6.

For S2, a sample of $N = 230$ Spotify users was recruited using the online recruiting service Prolific. All participants were native English speakers located in the US or UK; therefore, this study was conducted in English. To participate, an active Spotify account was necessary. 100 participants identified as male, 129 as female, and 1 person preferred not to tell. The mean age of the participants was 33.9 years ($SD = 11.4$). Participants were paid 2.80

GBP as compensation using the Prolific platform for the transaction. The study was preregistered under: https://doi.org/10.17605/OSF.IO/WE2ZP.

Both for S1 and S2 several other constructs were used in addition to SIPA and SIPA-P to assess construct validity. While Affinity for Technology Interaction (ATI, Franke et al., 2019) was surveyed in both, S1 and S2, following variables were additionally elicited in S2: Facets of System trustworthiness (FOST) scale (Trommler et al., 2018), Satisfaction & Usefulness (Van Der Laan et al., 1997) as well as Recommendation Experience (Hellmann et al., 2022). In addition to prior published scales, five items were devised to measure how confident the participant felt during the interaction with the recommendation algorithm (e.g., "I feel confident when creating a suitable station."). Finally, five items were added in S2 that were used to evaluate SIPA and SIPA-P and were developed for this purpose by multiple experts (e.g., "The questions on the previous page were confusing."). Both additional scales could be answered on a six-point Likert scale from 1 = completely disagree to 6 = completely agree.

## PROCEDURE

In both studies, participants were informed that participation is voluntary, that they will remain anonymous and that all collected data is handled based on the EU Data Protection Regulation. In S1, the participants answered the SIPA questionnaire as well as the ATI Scale. For S2 participants had to use their personal Spotify account and create a song radio based on a song of their choice and answered SIPA and the previously described variables. Their experience with SIPA and SIPA-P was elicited immediately after they answered the respective scale. An average of all items was calculated for evaluation. This process was repeated for the original SIPA scale after the participants answered questionnaires regarding perceived system trustworthiness, recommendation experience, system acceptance, and perceived confidence. To compare SIPA and SIPA-P, the completion time was measured in seconds.

## RESULTS

In both samples, a SIPA measurement using the SIPA scale was obtained only once (from here depicted as SIPA-1 for S1 and SIPA-2 for S2). In S2, the SIPA-P scale was completed additionally. In S1, SIPA-1 ratings were on average $M = 4.24$, ($SD = 1.11$). Descriptive data of all scale results from S2 can be found in Table 4. As we did not survey further variables relevant for the context of this study in S1, descriptive values for SIPA-1 can be found in detail Table 5. McDonald's omega ($\omega$) was used to assess reliability in addition to Cronbach's $\alpha$, since the latter is not well-suited for short scales (Dunn et al., 2014). Internal consistency of SIPA as well as SIPA-P was high (SIPA-S1: McDonald's $\omega = .93$, Cronbach's $\alpha = .93$; SIPA-S2: McDonald's $\omega = .91$, Cronbach's $\alpha = .91$; SIPA-P: McDonald's $\omega = .89$ Cronbach's $\alpha = .89$).

In S1, we found a significant but weak correlation between ATI and SIPA-1 ($r_S = .19$, $p < .001$). In contrast, in S2 there was no significant correlation between SIPA-2 and ATI ($r_S = .14$, $p = .070$). However, ATI and SIPA-P correlated significantly ($r_S = .33$, $p < .001$).

**Table 4.** Descriptive values for UX variable in in study 2 (S2).

| | *N* | *M* | *SD* |
|---|---|---|---|
| ATI | 230 | 3.38 | 1.02 |
| SIPA | 230 | 3.34 | 0.95 |
| SIPA-P | 230 | 3.31 | 0.98 |
| Recommendation Experience | 230 | 3.48 | 0.80 |
| Perceived Trustworthiness | 230 | 3.90 | 0.73 |
| Acceptance | 230 | 2.12 | 0.58 |
| Perceived Confidence | 230 | 4.03 | 0.81 |

**Table 5.** Descriptive values for SIPA-S1 ($N$ = 317), SIPA-S2 ($N$ = 230), and SIPA-P ($N$ = 230).

| | *M* | | | *SD* | | |
|---|---|---|---|---|---|---|
| | **SIPA-1** | **SIPA-2** | **SIPA-P** | **SIPA-1** | **SIPA-2** | **SIPA-P** |
| Item 01 | 4.54 | 3.14 | 3.61 | 1.28 | 1.20 | 1.28 |
| Item 02 | 4.26 | 3.02 | 3.21 | 1.33 | 1.13 | 1.12 |
| Item 03 | 4.36 | 3.54 | 3.28 | 1.29 | 1.15 | 1.30 |
| Item 04 | 4.21 | 3.43 | 3.31 | 1.35 | 1.17 | 1.23 |
| Item 05 | 4.07 | 3.38 | 3.27 | 1.27 | 1.13 | 1.25 |
| Item 06 | 3.98 | 3.51 | 3.21 | 1.29 | 1.15 | 1.16 |

*Note:* Kolmogorov-Smirnoff-Tests were significant for all items for SIPA-1, SIPA-2, and SIPA-P.

In S1, moderate correlations of SIPA-1 with both trust ($r_{\text{trust}} = .53$, $p = .003$) and usefulness $r_{\text{usefulness}} = .35$, $p = .003$) were statistically significant. In S2, neither acceptance nor the subscales usefulness and satisfaction correlated significantly with either the SIPA scale or SIPA-P scale ($p > .999$ for SIPA-P, $p = .282$ for SIPA-2 and satisfaction, $p = .156$ for SIPA-2 and usefulness). Trust (FOST) and confidence correlated moderately with SIPA-2 ($r_{\text{trust}} = .30$, p = .007 and $r_{\text{confidence}} = .26$, $p = .007$), as well as SIPA-P ($r_{\text{trust}} = .36$, p = .007 and $r_{\text{confidence}} = .25$, p = .007). Recommendation experience correlated moderately with SIPA-2 ($r_{\text{RecEx}} = .62$, $p = .007$) and highly with the SIPA-P scale ($r_{\text{usefulness}} = .73$, p = .007). Regarding differences in timing and evaluation of applicability between the SIPA-2 and SIPA-P, Student's t-Test for connected samples was conducted, showing small effect sizes which were not statistically significant (Cohen's $d_{\text{timing}} = -.05$, $p = .484$, Cohen's $d_{\text{evaluation}} = -.05$, $p = .479$). In addition, the correlation between SIPA-2 and SIPA-P was significant with $r = .52$ ($p < .001$).

We did not find a significant difference between participants' evaluation of the SIPA-P ($M = 4.19$, $SD = 1.00$) and the SIPA ($M = 4.24$, $SD = 0.97$), with $t(229) = .71$, $p = .479$, Cohen's $d = .05$. Additionally, participants time answering SIPA-P ($M = 40.4$, $SD = 34.3$) did not differ significantly from their time spent to answer SIPA scale ($M = 42.70$, $SD = 52.10$), with $t(229) = .70$, $p = .484$, Cohen's $d = .05$.

## DISCUSSION

The goal of the present research was to examine psychometric properties of the SIPA scale. The SIPA scale demonstrated high reliability in terms of internal consistency in both samples. Regarding construct validity, the SIPA scale demonstrated moderate to high correlations with trust but was not significantly correlated with usefulness and satisfaction. Correlations between SIPA and recommendation experience as well as perceived confidence were significant but not strong. The SIPA scale did not show any significant relationship with ATI, however, a significant correlation between SIPA-P and ATI was found. Also, recommendation experience correlated significantly with ATI. Comparing the SIPA scale with SIPA-P, no direct differences were found neither the completion time nor the scale evaluation differed significantly. Noteworthy, the correlation between both scales, SIPA and SIPA-P was only moderate.

## IMPLICATIONS

Based on the psychometric properties found, the scale seems suitable to measure user experience in Human-AI interaction. However, the analyzed data showed that when using such or similar measurement instruments, close attention must be paid to which constructs are included. Differences between the SIPA scale and the SIPA-P suggest that a person's general self-view may affect the evaluation of system interactions. The correlation of SIPA scale in plain language with ATI, as well as of recommendation experience with ATI, indicates an effect of the items' wording. In constructing the SIPA-P, e.g., several items were added that begin with "I..." and thus refer to the individual rather than the system. This might have caused the person's self-assessment to have a stronger influence on the responses than it did on the SIPA scale.

When using explanations in AI systems, false perceptions may arise. This can lead to an exaggerated understanding of the system, e.g., through explanations without diagnostic content (Chromik et al., 2021). In contrast to measurement methods from SA, such as SAGAT, SIPA offers the possibility to investigate whether users develop a false impression of their awareness. Thus, SIPA can be used, e.g., after explanations have been provided and SIPA scores can be compared with performance measures (Chromik et al., 2021). The ability to also consider subfacets can be advantageous in identifying the cause for increased values. e.g., it has been shown that an increase in information affects perceived transparency, but not necessarily perceived comprehensibility. This allows for an accurate assessment of experienced awareness and thus cognitive mechanisms through which interface changes impact the use of systems.

## LIMITATIONS AND FURTHER RESEARCH

When applying the SIPA scale based on the findings of this research, several limitations must be considered. At first, both samples only provided descriptive data on system use. There was a limited level of control (compared to, e.g., a lab experiment) and no experimental condition to actively manipulate

SIPA levels. Further studies need to apply guidelines developed to support information processing awareness and examine how this affects SIPA levels.

Furthermore, besides ATI, no personality measures were assessed in the samples. Potentially, Need for Cognition (Cacioppo & Petty, 1982) could also be connected to users' estimation of a system's ability to support them in understanding it. For example, users who tend to be skeptical when being confronted with AI systems may report lower SIPA levels, e.g., because they do not believe to have full access to the same information as the system. Therefore, while the SIPA scale is an excellent tool to shed light on user experience in automated information processing, further research is needed to examine how manipulations (e.g., explanations), as well as attitudes (see Park & Woo, 2022), can affect SIPA.

## CONCLUSION

The present research examines the Subjective Information Processing Awareness scale, a robust tool for assessing user experience with AI systems. Results from two different studies demonstrate high psychometric quality. Our results indicate that a less complex wording did not improve the SIPA scale. However, individual differences like ATI might impact user experience aspects like SIPA in automated systems, suggesting the importance of considering user characteristics in AI system design.

## REFERENCES


Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., … & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82–115.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Chen, J. Y. C., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science*, *19*(3), 259–282. https://doi.org/10.1080/1463922X.2017.1315750

Chiou, E. K., & Lee, J. D. (2023). Trusting automation: Designing for responsivity and resilience. *Human Factors*, *65*(1), 137–165. https://doi.org/10.1177/00187208211009995

Chromik, M., Eiband, M., Buchner, F., Krüger, A., & Butz, A. (2021). I Think I Get Your Point, AI! The Illusion of Explanatory Depth in Explainable AI. *26th International Conference on Intelligent User Interfaces*, 307–317. https://doi.org/10/gj5rvd

*Corona-Warn-App [Documentation]*. (2023). Retrieved from https://github.com/corona-warn-app/cwa-documentation.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making*, *9*(1), 4–32. https://doi.org/10.1177/1555343415572631

Endsley, M. R. (2018). Automation and Situation Awareness. In *Automation and human performance: Theory and applications* (pp. 163–181). CRC Press.

Fatemidokht, H., Rafsanjani, M. K., Gupta, B. B., & Hsu, C.-H. (2021). Efficient and secure routing protocol based on artificial intelligence algorithms with UAV-assisted for vehicular ad hoc networks in intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, *22*(7), 4757–4769. https://doi.org/10.1109/TITS.2020.3041746

Ferrario, A., Loi, M., & Viganò, E. (2020). In AI we trust incrementally: A multi-layer model of trust to analyze human-artificial intelligence interactions. *Philosophy & Technology*, *33*(3), 523–539. https://doi.org/10.1007/s13347-019-00378-3

Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. *International Journal of Human–Computer Interaction*, *35*(6), 456–467. https://doi.org/10.1080/10447318.2018.1456150

Hellmann, M., Hernandez-Bocanegra, D. C., & Ziegler, J. (2022). Development of an instrument for measuring users' perception of transparency in recommender systems. *IUI Workshops*, 156–165.

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2019). *Metrics for explainable AI: Challenges and prospects*. arXiv.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *IEEE Intelligent Systems*, *19*(06), 91–95. https://doi.org/10.1109/MIS.2004.74

Lorenz, B., Di Nocera, F., Röttger, S., & Parasuraman, R. (2001). The effects of level of automation on the out-of-the-loop unfamiliarity in a complex dynamic fault-management task during simulated spaceflight operations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *45*(2), 44–48. https://doi.org/10.1177/154193120104500209

Lyons, J. B., aldin Hamdan, I., & Vo, T. Q. (2023). Explanations and trust: What happens to trust when a robot partner does something unexpected? *Computers in Human Behavior*, *138*, 107473. https://doi.org/10.1016/j.chb.2022.107473

Park, J., & Woo, S. E. (2022). Who likes artificial intelligence? personality predictors of attitudes toward artificial intelligence. *The Journal of Psychology*, *156*(1), 68–94. https://doi.org/10.1080/00223980.2021.2012109

Popken, A., & Krems, J. F. (2011). Automation and situation awareness. In *The Handbook of Human-Machine Interaction*. CRC Press.

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, *28*(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

Sanneman, L., & Shah, J. A. (2022). The situation awareness framework for explainable AI (SAFE-AI) and human factors considerations for XAI systems. *International Journal of Human–Computer Interaction*, *38*(18–20), 1772–1788. https://doi.org/10.1080/10447318.2022.2081282

Schrills, T., & Franke, T. (2023). How do users experience traceability of AI systems? Examining subjective information processing awareness in automated insulin delivery (AID) systems. *ACM Transactions on Interactive Intelligent Systems*. https://doi.org/10.1145/3588594

Schrills, T., Kargl, S., Bickel, M., & Franke, T. (2022). Perceive, Understand & Predict—Empirical Indication for Facets in Subjective Information Processing Awareness. *CHI TRAIT '22*, 10. https://doi.org/10.31234/osf.io/3n95u

Trommler, D., Attig, C., & Franke, T. (2018). Trust in activity tracker measurement and its link to user acceptance. https://doi.org/10.18420/MUC2018-MCI-0361

Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*(1), 1–10. https://doi.org/10.1016/S0968-090X (96)00025-3

Zhang, Q., Lu, J., & Jin, Y. (2021). Artificial intelligence in recommender systems. *Complex & Intelligent Systems*, *7*(1), 439–457. https://doi.org/10.1007/s40747-020-00212-w