Intelligent Human Systems Integration (IHSI 2025), Vol. 160, 2025, 35–44
https://doi.org/10.54941/ahfe1005808

State-of-the-Art in Human-Centric Studies of AI-Enhanced Situational Awareness Within the Security Domain

Laura Salmela[1], Jussi Okkonen[2], and Roosa Heikkiniemi[2]

[1]VTT Technical Research Centre of Finland Ltd, Tekniikantie 21, 02150 Espoo, Finland
[2]Tampere University, Kalevantie 4, 33014 Tampereen yliopisto, Finland

ABSTRACT

Advanced situational awareness and decision-making systems in the security domain heavily build upon versatile combinations of different artificial intelligence models, potentially including their earlier versions. Often, the broad variety of implemented algorithms results in complex system architectures which may challenge human comprehension of expert users. System performance is frequently evaluated by different technical metrics against various data sets, such as model accuracy, precision, and recall. However, without any consideration of human-autonomy teaming or human-system interaction, the possibilities of executing comprehensive system assessments are likely to remain limited. This systematic review examines the current state-of-the-art in human-centric studies on situational awareness systems applying machine learning or artificial intelligence as key technologies. Our findings are based on up to 40 studies that were identified in our literature searches. This paper outlines the transition in research on the domain and current trends. It also discusses the research gap on human-centric approach.

Keywords: Situational awareness, Artificial intelligence, Security, Human-system interaction

INTRODUCTION

Versatile combinations of machine learning models (ML) form the core component of novel, advanced systems expected to enhance situational awareness (SA) and decision-making capabilities of security practitioners, such as military service personnel, police officers or border guards (e.g., Moore, Hebert, & Shaneman, 2018; Frontex, 2021). The algorithmic spectrum implemented within these systems is broad and often translates into complex system architectures that may challenge human comprehension of expert users. At single task level, however, the workings of an ML model might seem quite straightforward, understandable, and even familiar from other use cases external to the specific application area at hand (e.g., classifying an object from pictures or video frames and tracking its pattern-of-life or trajectory for a determined time period).

Frequently, the performance of novel situational awareness systems is evaluated with metrics addressing among others the accuracy, precision and recall of individual ML models against various data sets (UNICRI and INTERPOL, 2024). However, to comprehensively assess increases in situational awareness for human operators, a focus on algorithmic performance alone is likely to be insufficient, as no impact of human-system interaction is factored in. Despite autonomous or semi-autonomous decision-making may be vested within some of the developed systems, the human-in-the-loop concept generally persists as responsible artificial intelligence is highly pertinent in the security-critical domains, where authorities' decisions may have dire impacts on human lives (e.g., UNICRI and INTERPOL, 2024). This should be considered also in the perspective that in security domain there is requirement for human operation, thus man-out-the-loop is not even valid as system design requires human-in-the-loop or human-on-the-loop.

Complexity and time criticality characterize the environment in which military and other security personnel operate. According to Stanton et al. Boyd's OODA loop is one of the most widely used models aimed at depicting military decision-making (Stanton et al., 2008, 16). The OODA loop is a cyclical, internally linked model of military decision-making consisting of observe, orient, decide, and act. The model has been used to describe the importance of effective decision-making. Decision-making requires situational awareness, so it is essential to support as accurate situational awareness as possible, especially in the most time-critical tasks.

Situational awareness forms the starting point for the design of management systems. Situational awareness can be viewed as an individual, at the team or system level (Stanton et al, 2017). Simply put, situational awareness means understanding what is happening around (Stanton et al., 2017). Endsley (1995, 2017) divides situational awareness into three levels: perception, understanding and projection of the future. Situational awareness is a prerequisite for decision-making that leads to action that further shapes the state of the world and thereby produces feedback for perception. To the task, the environment and the individual related factors affect the formation of situational awareness. In practice, therefore, it depends on how complex A system is, how it presents information, and how interaction with it is like (cf. Stanton et al., 2012). The leverage of artificial intelligence in a military or security context are seen above all as the management of large amounts of data and the possibility to increase the autonomy of technical systems, the resulting benefits of which include, for example, better the tempo and influence of operations, as well as the possibility of decentralisation. In addition, AI can be used to model objects and events to be observed and predict the behaviour of objects based on learned information. However, the role of human operator and decision maker are still significant as in this context human-in-the loop or at least human-on-the-loop are required for ethical reasons.

In this systematic review, we examine the current state-of-the-art in human-centric studies on situational awareness systems that employ machine learning or artificial intelligence. The objectives are to identify predominant SA models and methods implemented for the assessment of situational awareness improvements, to investigate experimental designs or test set-ups of the situational awareness assessments, and to discover research gaps and suggest future research directions in the field. We draw our conclusions from close to 40 studies that were identified in our literature searches.

The paper is structured as follows. Review method describes the adopted methodology for a systematic review of current literature, while Results presents and discusses the review results by depicting identified articles and their characteristics. Conclusion concludes the paper.

REVIEW METHOD

We implemented our search query in August and September 2024 using the online search service *Andor* of the Tampere University Library providing access to over 350 databases[1]. We limited our search to four databases that were accessible through the same search interface: Academic Search Ultimate, Elsevier SD Completed Freedom Collection, IEEE Electronic Library, SpringerNature Journals and Web of Science. Using Andor's advanced search features, we also implemented additional filters to narrow search results prior to any relevance analysis: material type (*journals*), search language (*English*) and availability (*open access, available online, peer reviewed*).

We used three principles to define the search criteria which comprised different variants of selected keywords, including the primary concept of *situational awareness*, the specific technology of interest (*artificial intelligence*) and the specific application domain (*security/safety*). After a series of test searches, the following criteria were used for the final search query: *((“situational awareness” OR SA or “situation awareness”) AND (“artificial intelligence” OR ai OR a.i. OR “deep learning” OR machine-learning OR “neural networks”) AND (safety OR security OR military OR defence OR defense OR “law enforcement”))*.

The original search resulted in 724 potentially relevant journal articles. All search records were exported to Excel for screening against exclusion and inclusion criteria (Table 1) that were gradually developed after initial relevance examinations. Despite the aim of limiting our search to specific application domains, the initial query generated significant amounts of studies in other, unrelated fields (e.g., medical research). Journal and article titles were scanned to identify out-of-scope studies (e.g., research published in *Frontiers in bioengineering and biotechnology* or *Frontiers in neurorobotics*). Article abstracts and keywords were also screened, when the journal and article titles appeared too generic to infer relevance directly. With this approach, we eliminated 589 articles from the original result set.

We re-examined the titles, keywords, and abstracts of the remaining 135 articles, including also full-text screening. We further discarded 95 references on the basis that the studies addressed AI-model performance without any assessment of human-system interaction and its impact on situational awareness enhancements. The final screening also identified additional articles not meeting other inclusion criteria. These actions resulted in a final set of 39 articles for review.

[1] https://libguides.tuni.fi/az/databases

Table 1. Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Empirical, original research	Theoretical or non-empirical research (e.g., theoretical models, literature review, editorials, commentaries)
Full-text availability online	Full-text not accessible online
Publications in English language	Publications in other languages
Studies addressing decision-making in security or safety-critical settings	Studies addressing other application domains where AI-supported situational awareness systems are implemented (e.g., autonomous driving or transportation, business operations, consumer products, maintenance, process control, weather forecasting)
Studies reporting the measurement of situational awareness of individuals and teams or shared situational awareness (e.g., shared situational awareness between team members)	Studies reporting (technical) performance assessments that do not consider human component (e.g., performance evaluation of object detection algorithms)

We used qualitative data analysis software ATLAS.ti to extract key themes in the final reference set. As a first step, the full-text of each article (i.e., a pdf-file) was uploaded into ATLAS.ti. Then, we prompted the software to analyse the frequency of the occurrence of words and phrases in the article whole text, including abstract and key words. We clustered the articles by two periods (2017-2020 and 2021-2024) to examine development over years. With the exception of one article, 2017 represents the publication year of the oldest publications included in the sample, while 2024 the most recent. As noted, one article included in the sample was published in 2006 and is therefore not included in the comparison of word frequencies over time. In this context, the frequency refers to the number of times the words in each category were mentioned in the examined articles. For the examination of total frequencies, all publications are included.

Based on the automated code proposals provided by ATLAS.ti AI Coding Beta functionality, word categories were formed. Since the functionality proposes codes for the paragraphs in the texts and forms the results based on reading between the lines in these paragraphs, the exact word frequencies were counted using Python. The categories were formed to include both plural and singular forms of the words and different word forms were combined under the same category. For example, the category of “evaluation” contains frequencies for words “evaluation”, “evaluations”, “evaluate” and “evaluating”. Additionally, established categories contain associated terms or synonyms: for instance, the “drone technologies” category includes frequencies for the words “drone,” “swarm,” and “UAV”.

To define the most common categories, the occurrence of each article was examined across all the articles. The word frequencies were counted for those categories that appeared in at least 40% of the examined articles.

Ten categories with the largest frequencies in this group constitute the most common words.

RESULTS

In the following, we first provide an overview of the word frequencies, as they appear in the final article set covering publication years from 2017 to 2024. To indicate relevance of specific themes, we narrow our focus to words that are included in at least 40 percent of all examined articles (Figure 1), also examining evenly occurring words. Then, we present the word frequencies according to the two 4-year periods (2017-2020 and 2021-2024) and compare the results with each other. We conclude by investigating ascending and descending themes in the full article set to elaborate our analysis on potential shifts in thinking and research focus.

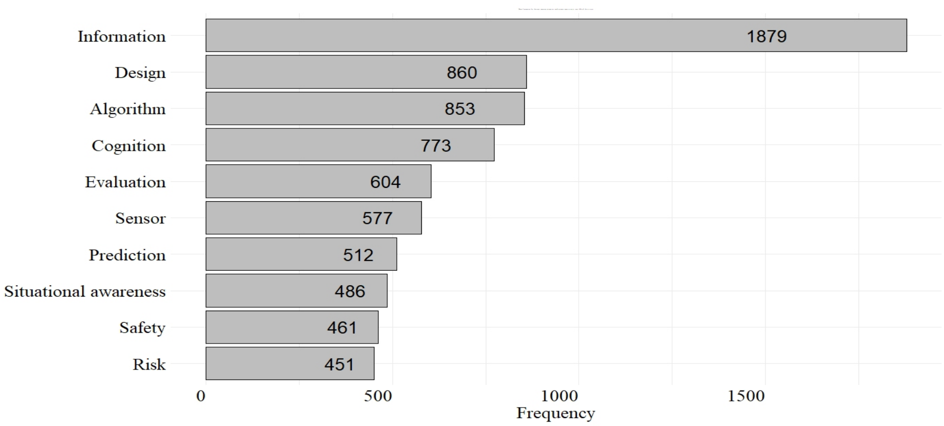

Figure 1: Word frequencies for the most common categories (each category comprising at least 40% of the articles).

Figure 1 presents the ten most common words with the words “information”, “algorithm”, “design”, “cognition” and “evaluation” being the most frequently used between 2017 and 2024 in the articles (i.e., representing the top 5 of used words). The results suggest that studies focusing on situational awareness and AI are showing interest in how technology especially contributes to information processes (incl. also risk assessment and decision-making) and human cognitive skills at different organisational levels and that of the individual. Information and cognition pair well and underline the notion of need to augment human capability in information related process. Design and algorithm additionally draw attention to need for technological advance to overcome human constraints. In other domains, the discussion on the role of technology could naturally be different. As visible from Figure 2 showing persistent word frequencies, cognition remains central to studies throughout the examined period from 2017 to 2024. Within the selected broad application domains of safety and security, particular attention has also been paid to sensor-generated data in support of these processes in contrast to other sources (e.g., open access or in-house databases) from which data can be queried and retrieved for analysis or mining purposes.

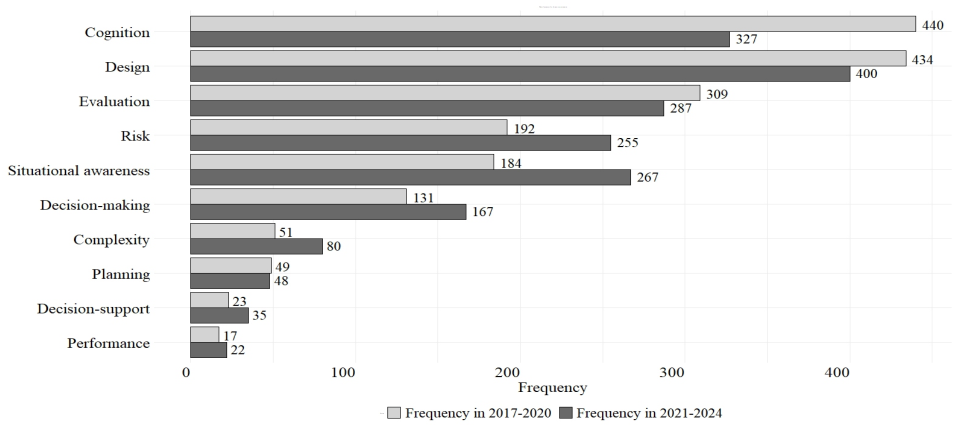

Figure 2: Word frequencies for the stable categories. The categories introduced in the graph were mentioned in at least three different articles.

Table 2 (excluding word frequencies from 2006 published article) shows the development of the top 10 most common words across all articles over time. There are notable frequency changes only with few words changing their relative position in the ranking order. For example, the use of the word "sensor" more than doubles in the second period (from 170 to 407). This indicates the importance of and the reliance on sensor-based data in generating situational awareness has become more recognised than earlier. In contrast, a significant decrease in the use of the word "prediction" might suggest that instead of using AI for *projecting* future events or scenarios for example based on historical data, current research investigates ways of enhancing real-time or near real-time analytical capabilities to assist situation *perception* and situation *understanding*.

Table 2. Top 10 words covering all articles according to selected time periods and frequency of occurrence.

Words (2017-2020)	Frequency	Words (2021-2024)	Frequency
Information	1085	Information	780
Algorithm	496	Sensor	407
Cognition	440	Design	400
Design	434	Algorithm	357
Prediction	339	Cognition	327
Evaluation	309	Safety	318
Risk	192	Evaluation	287
Situational awareness	184	Situational awareness	267
Sensor	170	Risk	255
Safety	141	Prediction	173

Figures 3 and 4 present the word frequencies of rising and declining categories. As noted earlier, the word "sensor" occurs significantly more often in the second period. Additionally, the word "autonomous" was used to close to five times more together with "human factor", "military", "adaptation" and "safety" also showing a visible increase in word frequency within the years 2021 to 2024. With regards to descending categories, changes in word use appear less drastic with the exceptions of the words "surveillance", "security", "communication" and "emergency" which have more than halved in frequency. However, for example in the case of the word "security", major changes in research emphasis are difficult infer as high word occurrences concentrated only to two years (2019 for the first time period and 2023 for the second) and to a limited number of individual publications.

Nonetheless, these upward and downward trends may signal that particularly in the military domain, increased use or interest towards autonomous, adaptive systems call for more research on human factors, including safety or human error considerations despite the systems' growing capabilities of operating independently with limited human interference. Significance of the human presence requires also considering ethics. This might coincide with one identified '*ironies of artificial intelligence*' arguing that "*the more capable the AI, the poorer people's self-adaptive behaviours for compensating for shortcomings*" (Endsley 2023, p. 1659).

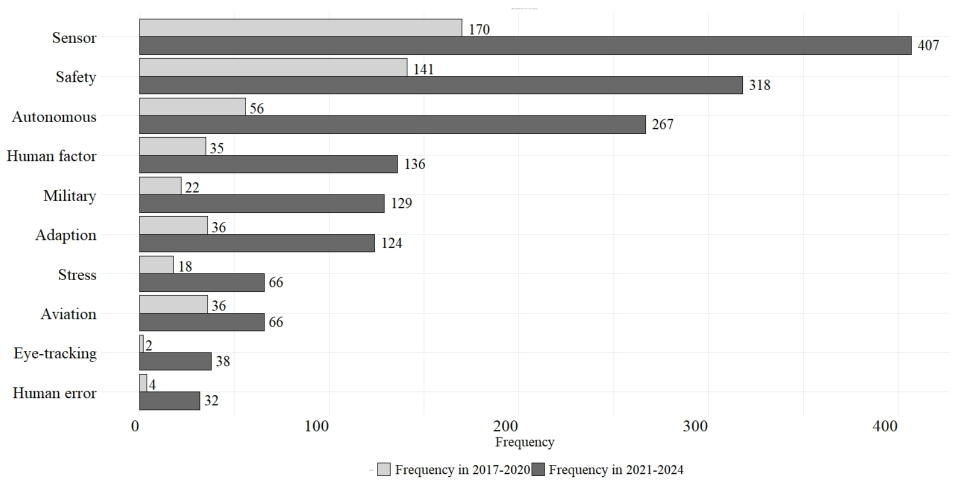

Figure 3: Word frequencies for the rising categories. The categories introduced in the graph were mentioned in at least three different articles.

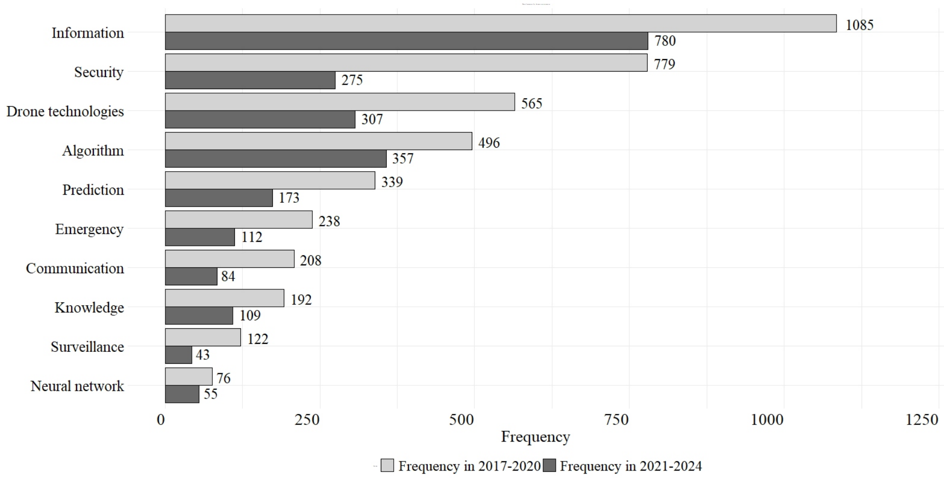

Figure 4: Word frequencies for the declining categories. The categories introduced in the graph were mentioned in at least three different articles.

CONCLUSION

This paper presents normative results for utilisation of technology in augmenting human perception and cognition. The findings are based on a systematic review of 39 articles from an initial 724 articles retrieved from four databases. The results highlight the assistive role of technology in maintaining situational awareness. Several studies discuss this as well as seek boundaries of human capability augmentation in this domain. Human perception, even if limited, has several advantages over the technology as well as technology can excel on different domains. Human perception is incapable of sensor fusion or beyond that, data-fusion. From a technology perspective, this has been reduced to somewhat simple integration problems yet from human perspective these are fundamental issues. Another significant topic is the role of human in human-autonomy teaming or human-technology integration, and this is often reduced to human-in-the-loop, human-on-the-loop, or human-out-the-loop distinction. As the analysis of key topics revealed the shift from technology to users, it reflects to contemporary discussion around the algorithms and AI. It has become everyday technology with numerous applications in almost all contexts, thus the focus could [finally] be shift from incremental development to user studies or experiments.

The development is significantly boosted by ubiquitous computing, mobile devices, and reduced cost of hardware. In military and security domain applications of narrow or more general AI are often integrated to certain vehicles or systems. The review did not show evidence on personal assistants such ones are familiar in consumer devices, applications, or services. However, the benefits of such technology, even in limited scale, are acknowledged and discussed. However, the domain calls for robust and explicable technology, so in near future most probably also situational awareness promoting technology is for specific use on certain use case, and with human-in-the-loop.

The applied review methodology did not come without limitations affecting the generalisability of our findings. Firstly, due to stringent exclusion and inclusion criteria, the sample size remained somewhat low considering the potential of the automated coding function provided by the analysis software. In future work, the sample size could be elaborated for example by analysing bibliographies of selected articles. This might enable to broaden the scope of examined literature by including a wider collection of both recent and less recent research. Additionally, examining correlation between words and their frequencies was considered beyond the limits of the present paper. However, such an approach could have substantiated our findings more. Finally, our analysis of ascending or descending word categories revealed only high-level themes. Investigating more specific research topics inside the selected application domains or human-centric approaches within those domains would have required more stratified content analysis.

ACKNOWLEDGMENT

The work of the primary author has received funding from the European Union's Horizon Europe Research and innovation programme under grant agreement No. 101073985.

DATA ACCESS STATEMENT

The final list of included articles for the literature review can be requested through the primary author.

REFERENCES

Baber, C., Harris, D., & Stanton, N. A. (2012). Modelling command and control: Event analysis of systemic teamwork. Ashgate Publishing, Ltd.

Battiste V., Lachter J., Brandt S., Alvarez A., Strybel T., Vu K. (2018). Human-Automation Teaming: Lessons Learned and Future Directions. In: Yamamoto S., Mori H. (eds) Human Interface and the Management of Information. Information in Applications and Services. HIMI 2018. Lecture Notes in Computer Science, vol. 10905. Springer, Cham.

Endsley, M. (1995). 'Toward a theory of situation awareness in dynamic systems', *Human Factors*, 37(1), pp. 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. (2017). 'From here to autonomy: lessons learned from human automation research', *Human Factors*, 59(1), pp. 5–27. https://doi.org/10.1177/0018720816681350

Endsley, M. (2023) 'Ironies of artificial intelligence', *Ergonomics*, 66(11), pp. 1656–1668. https://doi.org/10.1080/00140139.2023.2243404

Frontex – European Border and Coast Guard Agency. (2021) *Artificial intelligence-based capabilities for the European Border and Coast Guard*. Available at: https://www.frontex.europa.eu/assets/Publications/Research/Frontex_AI_Research_Study_2020_final_report.pdf (Accessed: 23 October 2024).

Moore, A., Hebert, M., Shaneman, S. (2018) The AI stack: a blueprint for developing and deploying artificial intelligence. In: Proceedings Volume 10635, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IX; 106350C (4 May 2018); Available at: https://doi.org/10.1117/12.2309483 (Accessed: 23 October 2024).

Stanton, N., Jenkins, D., Salmon, P., Walker, G., Revell, K., Rafferty, L. (2009). Digitising Command and Control: A Human Factors and Ergonomics Analysis of Mission Planning and Battlespace Management. https://doi.org/10.1201/9781315577241

UNICRI and INTERPOL. (Revised February 2024). *Toolkit for responsible AI innovation in law enforcement: Technical reference book*. Available at: https://unicri.it/sites/default/files/2024-02/07_Tech%20Ref_Book_Feb24.pdf (Accessed: 23 October 2024).