SUNDUS: A Human-Centered Framework for Fostering Human-AI Collaboration Through Transparency

Abduljaleel Hosawi[1], Richard Stone[2]
Dept. HCI, Iowa State University, Ames, USA[1]
Dept. CS, Jeddah University, Jeddah, KSA[1]
Dept. Industrial Engineering, Iowa State University, Ames, USA[2]

***Abstract*—Imagine a future where a security operator can understand a complex threat in fractions of a second without specialized training. Not only that, but the operator can instantly understand the logic behind a particular AI warning. This vision of AI-augmented cognition is the focus of the proposed SUNDUS framework. However, today, this promise faces a critical barrier: the user. The immense potential of AI is often rendered useless when its operator lacks the professional training necessary to interpret and implement its outputs in the real world. This context reveals a critical research gap: while collaborative human-AI systems significantly enhance performance, their efficacy remains fundamentally dependent on extensive operator training, as evidenced by the OMAR (Operator Machine Augmentation Resource) system. The current paper proposes the SUNDUS (System for Understanding, Navigating, and Decision-Making Under Uncertainty and Support) framework, a theoretical model designed through a Human-Computer Interaction (HCI) lens to enhance human-AI collaboration by making system design a substitute for formal training. Leveraging principles from AMID (Augmented Multisensory Interface Design) and Visual Representations of Meta-Information, SUNDUS employs enhanced transparency—via Natural Language Explanations (NLEs), confidence scores, and multisensory cues—to offload cognitive burden and increase intuitive understanding. We propose a comparative experimental methodology to validate SUNDUS against OMAR, hypothesizing that SUNDUS will yield significantly higher decision-making accuracy and appropriately calibrated trust alongside a lower cognitive load in untrained users. The key implication is a scalable, human-centric design blueprint that shifts the burden of adaptation from the operator to the AI system, unlocking the full potential of augmented cognition.**

***Keywords*—SUNDUS; human-AI collaboration; Human-Computer Interaction (HCI); transparency; multisensory display; cognitive load; decision-making; trust; training; OMAR**

I. Introduction

A. Background and the Broad Aim

The development of artificial intelligence is rapidly evolving from a focus on full automation to a more nuanced approach centered on human-AI collaboration [1]. A notable feature of this trend is the growing recognition of AI as a source of effective augmentation of human operator capabilities [2]. It has previously been viewed as a mere tool for humans or a substitute for them. AI systems outperform humans in their ability to process massive amounts of data, identify patterns, and perform repetitive tasks with a speed and efficiency beyond human capabilities [3]. This superiority has created a significant opportunity for human operators to focus on higher-level tasks such as those requiring moral reasoning, wisdom, and contextual understanding—intelligence traits unique to humans and superior to AI. The synergy between human operators and AI allows them to leverage the strengths and uniqueness of both.

The OMAR system, designed to support the operator in their task, is a clear example of this collaborative approach [3]. The goal was not to replace one party, but rather to improve the performance of the human CCTV operator in crowd tracking tasks by utilizing AI capabilities in several aspects, including the following:

1) Training: The study's results on the Operator Machine Augmentation Resource (OMAR) system clearly indicate the importance of comprehensive operator training as a key element for success [3]. Through synergistic work, AI capabilities are leveraged to increase efficiency and speed. However, the OMAR study found that AI tends to increase errors in untrained groups, highlighting the essential role of human expertise gained through training. Training increases operator proficiency in validating, interpreting, and utilizing AI outputs, while also reducing potential errors and enhancing trust as the OMAR system [3] has shown.

2) Decision-making tools: AI is a key component of the OMAR system's operating team. It employs a computer vision model that helps detect and report anomalous behavior during human crowd movements. Instead of an operator constantly scanning crowd flows for hours on end, the OMAR system initially alerts to potentially anomalous behavior. The operator examines these alerts and analyzes their behavior to make appropriate decisions, as shown in Fig. 1.

3) Reducing cognitive load: Human operators are subject to significant cognitive load while performing their duties, scanning scenes on surveillance screens. Artificial intelligence (AI) acts as a powerful assistant, relieving the operator (especially trained ones) of this mental effort by detecting persistent and anomalous patterns, as illustrated in Fig. 2. This feature allows the human operator to remain more alert and focused, especially when detecting situations requiring human skill. Ultimately, shared monitoring responsibilities improve the efficiency and accuracy of operations and enhance trust.

Therefore, the OMAR framework provides strong empirical evidence that human-AI teamwork outperforms either humans

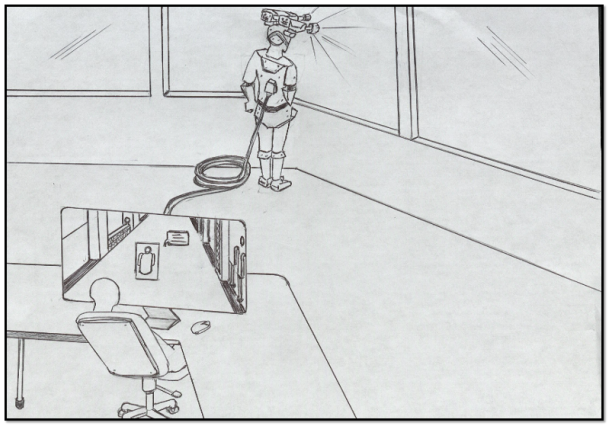

Fig. 1. OMAR system illustration, showing abnormal behavior detected and a short message displayed to the operator.

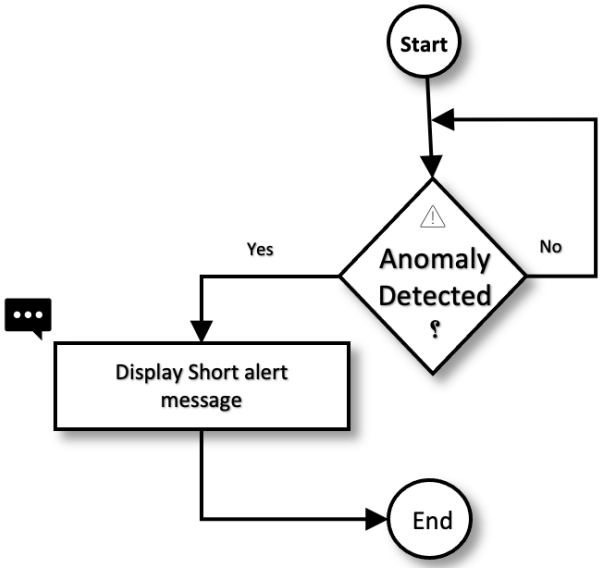

Fig. 2. Information flows from the OMAR system to the operator, helping to reduce cognitive load.

or AI alone. This superiority can be further enhanced if training is carefully combined with technological support.

B. The Research Gap

The rapid increase in artificial intelligence (AI) use in various fields, particularly high-sensitivity ones such as safety, has revealed a new paradigm of human-AI collaboration. The proliferation of this paradigm has exposed a critical research gap. While AI systems are designed to enhance the capabilities of human operators, human operators often lack the specialized training necessary to utilize the technology effectively[4]. The OMAR system provides a leading example of this gap, demonstrating conclusively that operator performance is significantly higher with comprehensive training. In reality, in many contexts, most people do not receive this type of specialized, in-depth training to work with AI systems for various reasons, such as time constraints, costs, technological acceleration, and more. This gap results in significant missed opportunities, as operators cannot fully leverage AI's capabilities without a solid and sound foundation of expertise, which often comes from training. This problem has been observed in the OMAR system. While the system's alerts are technically valid, they are in fact vague and can increase confusion for human operators (especially untrained ones). When combined with the difficulty of verifying their justification, the alerts are often ignored, which can lead to unforeseen risks. Despite OMAR's significant role in reducing mental effort and assisting in decision-making, this role remains limited and insufficiently autonomous. If it continues in this limited form, this could lead to a breakdown in trust and a decline in integration between the operator and artificial intelligence. Ultimately, the system's effectiveness remains fundamentally dependent on operator training, which raises the question of whether a different design approach could mitigate these issues. These gaps lead us to the central question of this research: Can system design replace formal operator training through transparency and multisensory information? This paper's key novelty and central hypothesis directly address this by proposing the SUNDUS framework, which demonstrates that intelligent system design can effectively substitute the cognitive benefits typically acquired only through intensive operator training.

It is critical to note that this work serves as a theoretical framework and research proposal, outlining a novel conceptual design and the testable hypotheses derived from it. As such, this paper does not present empirical data; instead, the traditional results section is replaced by a detailed Future Work section outlining the proposed comparative experiment necessary for validating the SUNDUS framework. The remainder of this paper is organized as follows: Section II establishes the contextual background, detailing the limitations of the OMAR system and reviewing the theoretical foundations (AMID and Meta-Information) that underpin this design. Section III presents the full SUNDUS framework, detailing its components and features. Section IV provides a comprehensive discussion of the framework's theoretical promise and practical implications. Finally, Section V outlines the conclusions and the specific experiment planned for future validation.

II. Contextual Background and Related Work

A. Human Information Processing and Memory

1) Perception and attention: Perception and attention are among the first steps in human information processing. In the OMAR system, an operator's awareness depends on their ability to see the "Anomaly Detected" message on the screen. Once the operator recognizes the message, they focus on some of its details and characteristics, which means paying attention to those details in preparation for an appropriate decision. Focusing and attention require a significant amount of mental effort. Without adequate training, the operator may struggle, in particular, to interpret the content of the anomalous behaviors detected by the system. The situation becomes even more acute if the operator suffers visual fatigue or distraction. The OMAR system supports the operator's ability to monitor the screen and understand the system's alert, accompanied by a brief text message indicating an anomaly.

In contrast, the proposed SUNDUS system is designed to address the limitations of human perception and help directly

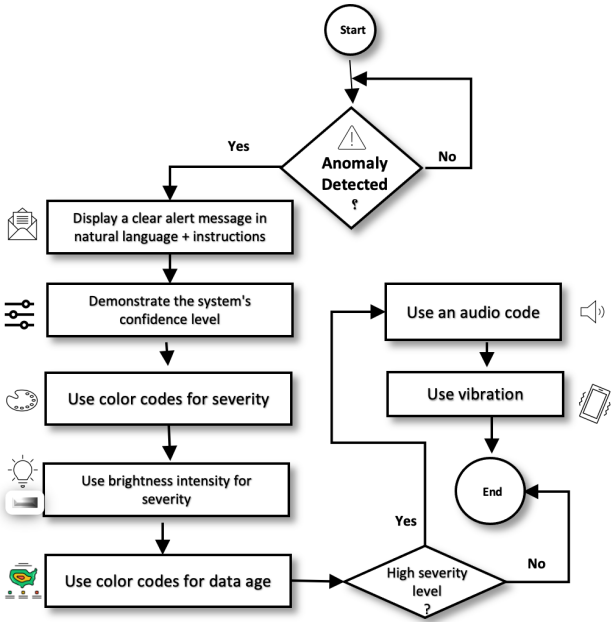

Fig. 3. Multisensory information flows from the SUNDUS system to the operator.

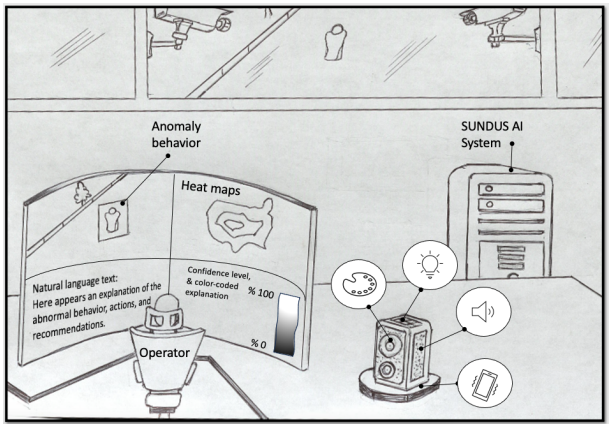

Fig. 4. A conceptual view of the SUNDUS system and its most prominent features.

increase attention. It utilizes multiple sensory channels to communicate event data, particularly through underutilized channels such as auditory and tactile, as the AMID principles indicate [5]. Audio channels can be used with distinct tones to indicate specific descriptive states to enhance alerting to high-priority events. Similarly, vibration can draw the operator's attention to critical events even when other channels (visual, auditory) are busy or overloaded. By leveraging multiple sensory capabilities, the SUNDUS system will likely overcome the bottleneck of using a single sensory channel. Additionally, the SUNDUS system utilizes visual representations of meta-information, employing colors and shapes to convey important details at a glance. As in [6], this shift can be achieved through a flashing color gradient bar indicating severity, or highly saturated heat maps.

Furthermore, incorporating visual representations of meta-information into SUNDUS can support decision-making, particularly in time-sensitive or high-risk environments. Some examples of this application in SUNDUS could be adding a flashing color gradient bar or highly saturated heat maps (see Fig. 3 and Fig. 4). These elements can help draw the operator's attention to the most relevant information about the event and its context, while assisting in assessing finer details such as severity, certainty, or age of the information through color saturation. Implementing these features is expected to reduce the need to search for anomaly details and speed up the response process, in addition to saving the mental effort required to integrate multiple pieces of information to obtain a clear picture and make a decision.

2) Working memory and cognitive load: Working memory is like a temporary repository that an operator must access to gather fragments of information about anomalies for analysis and understanding in preparation for a response. One challenge that can occur is when this repository (working memory) becomes full; operator performance declines sharply. The OMAR system is designed to flag anomalies on the monitor screen and display a brief message indicating that an anomaly has been identified. The operator then examines the details of the anomaly by revisiting working memory, assessing the event, and making the appropriate decision. This process undoubtedly requires significant mental effort, especially given the continuous flow of images and the potential for additional detection, which increases the chance of errors that could lead to risks or a decline in overall performance. An operator's lack of specialized training increases the chance of problems and errors.

On the contrary, the SUNDUS system is designed to reduce the load on the operator's working memory. The operator does not need to use working memory to try to remember and gather details about anomalous behaviors. Instead, the system provides an immediate visual summary. For example, meta-information qualifiers (e.g., uncertainty, age, source reliability) may be presented directly on the screen. This design means the system performs complex cognitive tasks, allowing the operator to focus on more important tasks such as decision-making.

3) Long-term memory and the role of training: To help operators remember important details necessary for their job performance, such as regulations, processes, recommended procedures, and others, they must be stored in long-term memory and reinforced over time. Considering how the OMAR system works, we find that a mental model is created in primary long-term memory and then reinforced with training. Therefore, the OMAR paper suggests that training is essential for effective synergy between human and AI teams.

On the other hand, the SUNDUS model proposed in the current framework paper seeks to be more intelligent by providing an external mental model based on its unique design. While OMAR relies on training to accumulate what the operator needs to remember in long-term memory, SUNDUS displays this information using color codes to directly represent deep, detailed data without requiring the mental effort of retrieving it from long-term memory, or even extensive prior training. Imagine representing the severity of an abnormal behavior with a color-graded bar, or the system's confidence in certain information with different levels of accuracy, alongside

clear messages in natural language with recommended actions. With this design, SUNDUS will likely support long-term memory performance by saving the effort and time required to store and retrieve data. Furthermore, by providing assessments and recommendations in natural language, the system will improve the overall team performance by saving effort and time and supporting decision-making, ultimately enhancing trust between the operator and the system.

In short, the SUNDUS system design is unique because it does not require prior human cognitive capabilities to be built into the operator through in-depth training, as is the case with the OMAR system. Instead, it aims to achieve high team performance even without extensive training.

4) Training as an enhancement method: Training is one of the fundamental methods for enhancing human capabilities and improving the quality of human-AI interaction. The importance of training is emphasized with complex systems, where the role of training goes beyond simply transferring knowledge to the human operator. Typically, training hones the operator's cognitive skills and develops their capabilities, overcoming weaknesses and developing strengths. Training is also an effective means of helping form mental models of the system's work. The operator can gain a clearer picture of the AI system by delving into its operational logic, system design, objectives, and inherent limitations through training. In addition, training can play a fundamental role in how humans receive and critically evaluate information and then act on the results. All of the previous contribute to reducing errors and problems, improving performance, and transforming the operator from a passive observer to an active partner in the synergistic process.

The OMAR framework provides empirical evidence that confirms the power of training as an improvement tool, as the results confirm the relationship between operator performance and training. The study demonstrated that operator capabilities in the OMAR system improve with training. The research showed that participants who received specialized training significantly and consistently outperformed their untrained counterparts. This result was demonstrated by a significant reduction in errors and their ability to detect them, along with a significant increase in identification accuracy. On the other hand, researchers observed that the use of AI caused an increase in errors in the untrained groups, in contrast to the groups that received training.

Furthermore, training allowed operators to leverage better AI capabilities, including the system's speed and efficiency in identifying anomalies. The main finding of the research confirms that training is beneficial for improving human performance and crucial for enabling effective synergy between humans and artificial intelligence. It also contributes significantly to reducing errors. Thus, the OMAR system clearly demonstrates the role of training in enhancing human operator capabilities by providing them with the cognitive tools necessary to work with artificial intelligence, which promotes high performance levels and harmonizes human and machine capabilities.

5) The need for a new approach: Although training is important to enhance operator capabilities, as demonstrated by the OMAR system, this approach is not always feasible or available. In many cases, especially in the real world,

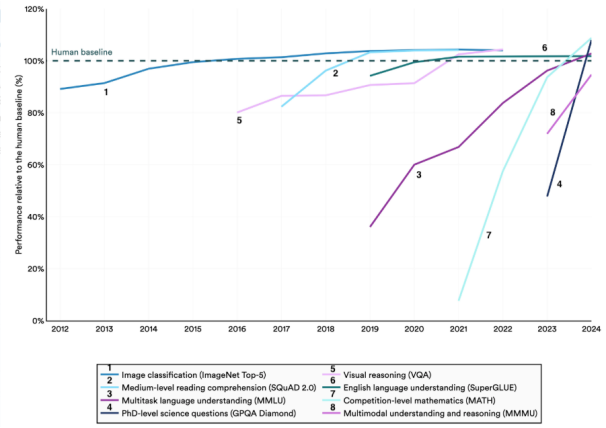

Fig. 5. The continuous evolution of AI performance compared to human capabilities encourages its applications. Reproduced from [2].

most people do not receive the in-depth training necessary to operate AI systems effectively for various reasons. Reasons may include financial costs, the sheer number of potential users, and time pressures, especially in a rapidly evolving technological environment, which leads to a significant gap between the advanced capabilities of AI and the user's ability to leverage them. This previous fact leads us to the central question of this research: How can we enable the operator to fully leverage the capabilities of an AI system with minimal, or even no, in-depth training? This means exploring a new mechanism or approach to designing human-like AIs, specifically those systems that enable effective collaboration with the operator—not through extensive human training—but through their inherent characteristics. Therefore, we propose that the solution lies in an AI system designed to be a transparent and intuitive partner that supports a practical and comprehensive partnership even without prior operator training. Such design can be achieved by leveraging principles of human perception. The proposed new approach is likely to help shift the burden of adaptation from the operator to the system, helping to make the full benefits of AI available to a much wider audience.

B. The OMAR System and the Training Paradox

1) Theoretical foundation: Rapid development of AI systems has created a new paradigm in human-computer interaction, driven by collaborative work and integration resulting from capitalizing on the strengths and uniqueness of both humans and AI, which is considered a means of enhancing the capabilities of the human operator, as shown in Fig. 5. However, the human operator may suffer fatigue due to their limited capabilities compared to the massive amounts of data AI can process and generate. This problem is exacerbated by the lack of or inadequate training to operate these systems. The work in this research is based on two main theoretical frameworks that form the basis for designing AI systems that more effectively enhance human capabilities while addressing fundamental gaps, especially in the absence or insufficient provision of intensive training.

In Fig. 6 and Fig. 7, the growing interest in AI is demonstrated in many fields through regulations and businesses. In 2024, US federal agencies issued nearly 60 regulations related

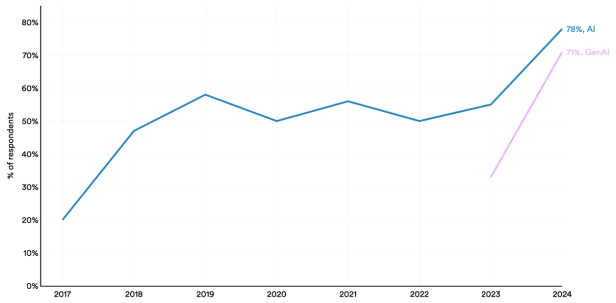

Fig. 6. Respondents who confirmed the use of AI in their organizations, 2017–2024. Adapted from [2].

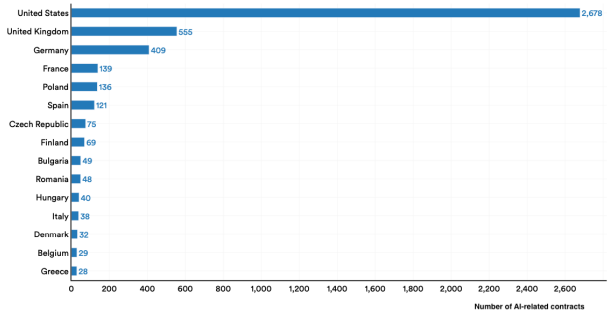

Fig. 7. Growth in the number of AI-related contracts in selected countries, 2013–2023. Adapted from [2].

to artificial intelligence, which is more than double compared to 2023. Globally, there was rapid growth in several countries by 21.3% in AI-related investments between 2016 and recent years [1]. Canada, India, and France pledged investments of 2.4, 1.25 billion, and €109 billion, respectively. Meanwhile, Saudi Arabia announced a $100 billion initiative as part of its 2023 vision.

2) The deep learning black box: The accelerating pace of human-artificial intelligence synergies is attributable to deep learning (DL) architectures, particularly Convolutional Neural Networks (CNNs) and vision transformers in monitoring systems. These technologies are characterized by their high capacity to provide unparalleled speed and accuracy in pattern recognition (as in the components of the OMAR computer vision system). Both DLs and CNNs inherently function as algorithmic black boxes [7]. The ambiguity arises from the vast number of non-linear transformations within the complex network used in the decision-making process. These systems not only create a lack of transparency or pose a technical challenge but also present a real problem for untrained human operators and raise ethical concerns that may affect human trust in them, especially in high-risk domains, leading to inappropriate reliance [8], [9], [10]. The problem is conceivable if the operator cannot understand why the system is issuing an alert or does not clarify the degree of trust in the information provided by the system.

3) Existing human-AI interaction models: Several effective models have been developed to gain a deeper understanding of human-AI interaction, on the one hand, and in response to the need for a comprehensive framework to explore and improve this type of collaboration, on the other [11]. However, despite

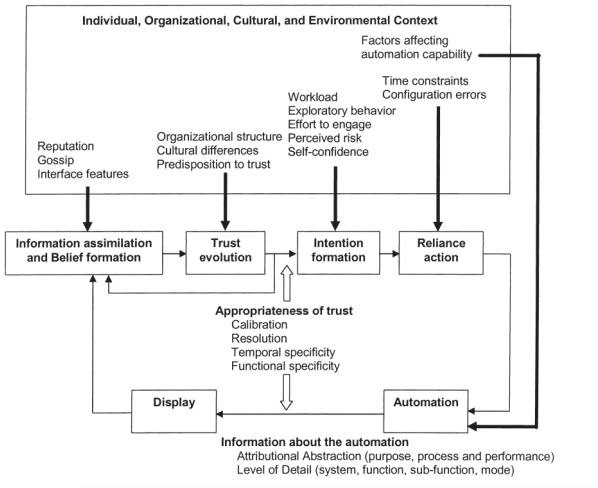

Fig. 8. A conceptual model illustrating the elements of trust and their impact on dependence in organizations. Adapted from [10].

their importance, many do not fully address the challenges when an operator lacks specialized and sufficient training to handle AI systems.

One prominent school of thought focuses on trust in automation [10]. This model assumes that a human operator's willingness to rely on AI depends on three factors: the system's perceived reliability (its capability, safety, and goodwill), the operator's innate propensity to trust, and situational factors as shown in Fig. 8. Despite its significant influence, this model often implicitly assumes that the operator has sufficient knowledge and skills—usually acquired through training—to calibrate their trust in the system properly. It is certain that without a basic understanding of AI and the optimal way to handle it, an untrained operator may exhibit risky behavior, ranging from over-reliance on the system to complete abandonment due to confusion or lack of trust. Both over- and under-reliance on AI can lead to suboptimal outcomes. In this context, the trust model clearly describes how trust is formed. However, it does not provide a mechanism for establishing appropriate trust when human knowledge base or practical skills are lacking.

4) XAI and multisensory solutions: The efforts to solve the black-box problem of deep learning have led to the emergence of the field of Explainable Artificial Intelligence (XAI), which aims to provide human-comprehensible explanations for the decisions made by AI systems [12], [13]. To achieve these solutions, various techniques are employed, such as feature importance maps (similar to prominence or heat maps) or counterfactual explanations [14]. While studies in XAI have often focused on visual interpretations, recent work aligns with the principles of the AMID framework by exploring multisensory media to represent complex deep learning outputs [15], [16]. The importance of this shift is underscored by the understanding that overloading the visual channel with data visualizations can quickly lead to increased cognitive load and diminish the benefits of transparency [17], [18]. Therefore, current trends [19] involve designing systems that leverage a combination of visual prominence tools and underutilized sensory channels (haptic and auditory) to convey the confidence

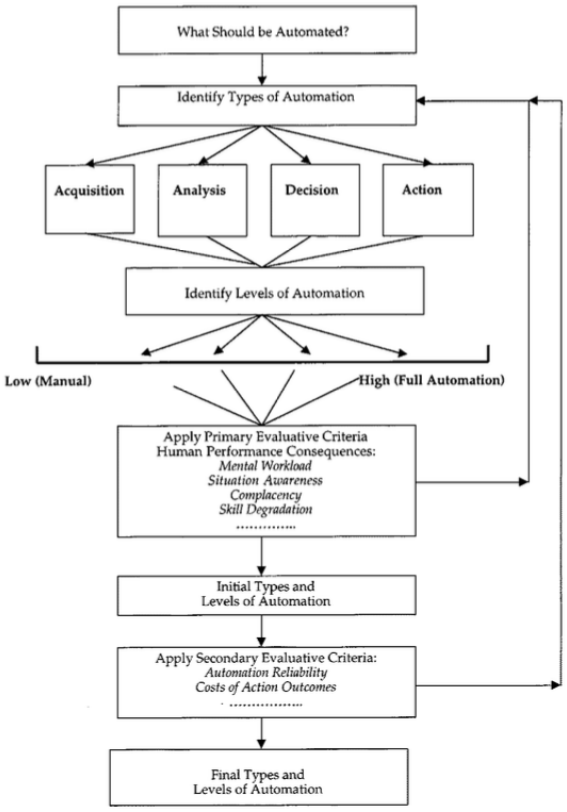

Fig. 9. Application of the automation types and levels model. Adapted form [20].

and rationale of AI effectively.

Other models focus on the levels of automation when dealing with AI systems, establishing a categorization ranging from complete human control to complete AI autonomy [20]. While this approach effectively classifies different system designs, it lacks the flexibility to adapt to changing contexts. It is a model that can be likened to the rungs of a ladder, as the operator's work environment may require dynamics depending on the context. For example, different tasks may require varying levels of automation for the same operator, and automation may need to be calibrated based on the operator's cognitive abilities, especially when considering the availability and lack of training (see Fig. 9). These contexts can make it challenging to know when to intervene or when to let the AI take over. The previous challenge confirms that an operator not equipped with the necessary skills and knowledge lacks an acquired mental model that enables them to calibrate appropriately based on the AI's capabilities and limitations. Furthermore, lacking training may mean being unprepared to make critical decisions in real time.

In study [21], the IICCM framework is introduced, a comprehensive system to support operator tasks in enhancing safety. IICCM integrates artificial intelligence, computer vision, and the Internet of Things to manage large crowds. The IICCM framework supports a leading suite of services, including early emergency detection using GANs and ViT networks; real-time evacuation route recommendations via mobile applications and local positioning systems (GPS); and

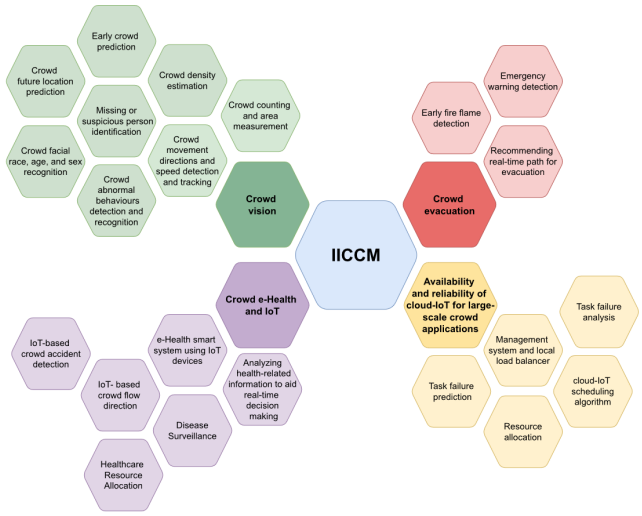

Fig. 10. Components and services of the proposed Intelligent Integrated Crowd Control and Management (IICCM) framework. Adapted from [21].

the integration of behavioral models to guide safe evacuation (see Fig. 10). However, IICCM does not eliminate the need for in-depth operator training to perform their tasks effectively.

Recent research highlights a shift toward exploring frameworks that capture modern operator-AI collaboration's complexity. One example is "shared cognitive systems," which views the operator and AI as a single entity capable of performing cognitive lab work [22]. This approach focuses on designing a "cognitive work environment" that supports reducing the cognitive load and fatigue experienced by the human operator. Additionally, recent frameworks have highlighted the importance of placing the human at the center of attention from the outset, while maintaining the transparency and interpretability essential to the human partner's needs. Another study by the AI Partnership [4] developed a comprehensive framework that addresses 36 questions to help design collaborative work by analyzing and understanding the nuances of various forms of collaboration between operators and AI. The analysis helps understand the factors influencing collaborative work contexts, such as the specific task and operator expertise. Human-centered AI (HCAI) has recently been formalized to focus on human needs, values, and capabilities when designing AI systems [1]. Upon reflection, all of these frameworks, including SUNDUS, address the limitations of traditional models by calling for AI systems that enhance, rather than replace, human capabilities, and for future systems to be transparent, reduce cognitive load, and ethically aligned with human values. However, they differ in their focus, level of abstraction, and practical application. While the framework of "Human-Centered Artificial Intelligence" (HCAI) a cognitive systems serve as philosophical schools of thought that justify the need for a framework like SUNDUS, they justify the underlying reasons behind their philosophy. SUNDUS provides the "how" by translating these philosophies into a reality with a testable blueprint with specific properties. These theoretical frameworks provide a strong foundation for the research addressed in this paper.

Ultimately, the limitations of modern basic models and the

opportunities they reveal a significant gap. There is an urgent need for a framework that provides a practical and executable blueprint for designing an AI system capable of acting as a transparent and intuitive partner, effectively augmenting human capabilities in situations where formal operator training is limited or absent. This is the gap that the SUNDUS framework aims to fill.

C. Foundational Theories (AMID, Meta-Information)

SUNDUS has two solid foundations. The first is the Augmented Multisensory Interface Design (AMID) framework [5]. The AMID framework states that effective augmentation is not limited to adding more visual information to a screen. Instead, it includes designing a system that strategically utilizes all available human sensory channels—visual, auditory, and tactile—to communicate information and reduce cognitive load. One of the core principles of AMID is to enhance an underutilized sensory channel for information distribution, freeing up a primary, often overloaded, channel (such as vision). This principle may strongly and directly support the goal of this study, which is to improve human performance without training. Multisensory design can provide intuitive, non-visual cues that an operator can understand and act upon immediately without prior instruction or training.

Second, the research on visual representations of meta-information, as [6] investigated. Based on this work, the effectiveness of a visual representation depends on its ability to communicate data parameters (meta-information) such as uncertainty, data age (freshness), and source quality. Notably, the research found that the perceived importance of information is the most important factor in how humans interpret visual cues such as color, saturation, and transparency. These findings support the possibility of designing AI systems capable of communicating complex and profound concepts, such as "the reliability of an AI data source" and "the freshness and age of the data," by manipulating visual attributes. Therefore, these features can be leveraged to produce a transparent, intuitive system that allows an untrained operator to deeply understand the reliability of the alerts displayed by the system on the screen simply by looking at them.

III. The SUNDUS Framework: A Human-Centric Model

The SUNDUS (System for Understanding, Navigating, and Decision-Making Under Uncertainty and Support) framework is a conceptual model designed to bridge the gap between human operators and advanced AI capabilities, especially when human operators lack specialized training. While the OMAR system requires training as a critical factor for improved performance, SUNDUS's principle fundamentally differs: Its unique contribution is the demonstration of how rich, transparent system design can directly replace the need for an operator's specialized, trained-in mental model, thereby allowing SUNDUS to outperform OMAR through intuitive strategic operation.

1) Core principles: The basic principle of the SUNDUS system is the strategic presentation of information necessary to achieve system transparency and to utilize multiple senses within the operator. The primary goal is to reduce the cognitive load by transferring functions to the system, transforming the burden of ambiguous system alerts into intuitive, information-rich interactions. SUNDUS achieves this goal by transferring cognitive functions—such as interpreting system confidence, prioritizing alerts, and recalling procedural knowledge—to the system, which can be directly displayed through a human-centric interface. Fig. 11 illustrates this core principle. These cognitive functions were previously only acquired through in-depth training.

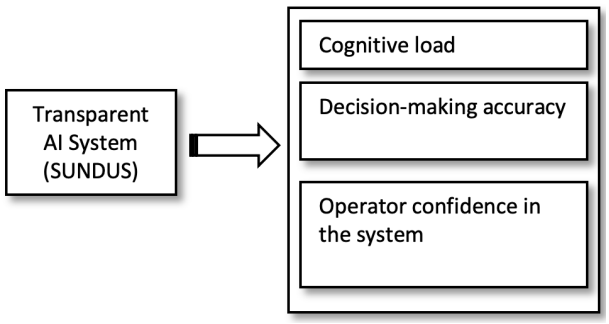

Fig. 11. Basic principle of the SUNDUS system.

2) Framework components and features: To build a system that can collaborate with a human operator in a synergistic environment without the need for extensive training, as envisioned in the SUNDUS framework, several new features must be provided that support conditions that eliminate the need for costly training, such as enhanced transparency, cognitive offloading, and improved performance [23],[24]. These features include:

3) Color-coded salience: The first feature of the SUNDUS system is the use of colors to indicate the severity level of an alert. For example, red can indicate a high-priority threat, while yellow can indicate a lower priority.

4) Representing trust with saturation or brightness level: It leverages humans' mental association to link color intensity to confidence levels. Therefore, the SUNDUS system uses high color saturation or brightness to indicate the AI's confidence level in an alert [25]. For example, high red saturation (or brightness) indicates high confidence, while low brightness or saturation indicates low confidence. This feature allows an operator to intuitively assess the reliability of an alert, even without prior training.

5) Color-graded bars and heatmaps: SUNDUS can provide a color-graded bar on the video feed to indicate the depth of information. This visual representation can help determine where the AI focuses, which areas are changing, or which locations have the highest confidence, giving the operator a comprehensive visual understanding of the situation. Similarly, a visual heatmap overlay can be used.

6) Multisensory cues: To enhance operator attention without overwhelming the visual channel (the most frequently used), the SUNDUS framework integrates multisensory cues. This approach builds directly on the central principle of the AMID paper, which is to utilize underutilized sensory modalities to improve performance. One scenario this paper outlines is using meaningful audio cues to differentiate between routine

and critical situations. A low-pitched ambient tone could signal a normal alert, while a high-pitched, sharp tone could signal a critical threat requiring immediate action. The goal of this strategy is to relieve the load on the visual channel, enabling the operator to respond even if this channel is occupied by something else.

7) Haptic alerts: Critical alerts of utmost importance require a more effective method for communicating alert data. Therefore, the SUNDUS system employs a haptic signal (such as a vibration on a terminal or wearable device) [25]. As demonstrated in the AMID method, this strategy is effective, as haptics are ideal for alerts and are associated with manual reaction speed. This association makes them a highly effective channel, especially since they are often underutilized for conveying urgent information.

8) Context-aware interpretation and clear guidance: One of the most important features of the SUNDUS system, which plays a significant and direct role in compensating for the lack of training, is its ability to provide context-aware guidance. This feature transforms the system from a passive alert tool to an effective collaborator through the following methods:

a) Natural Language Explanations (NLEs): While the alert is important in the Omar system, it is vague and does not explain the alerts. SUNDUS addresses this issue by supporting the operator with a concise natural language explanation (NLE) for reporting a particular incident. For example, instead of displaying a short message stating "Anomaly detected," the system could provide more detail, such as: "Alert: Anomaly detected. There is a sudden gathering of more than 10 individuals in a restricted area. This event does not conform to typical crowd flow patterns." An explanation like this provides a rationale typically learned through operator training. A message like this and its clear explanation enhance immediate understanding of the situation.

b) Actionable recommendations for the operator: SUNDUS will provide "suggestions on how the operator is recommended to act in specific situations." These recommendations, which can be as quick and straightforward as "Recommendation: Contact a security officer for verification," can provide the operator with a clear and immediate course of action. This feature supports the operator's decision-making process by providing a procedural framework that compensates for a lack of formal training or acquired experience.

IV. Discussion: Rationale and Implications

The specific features of the SUNDUS framework are deliberately designed to create a new paradigm in human-AI collaboration. This section discusses the rationale for the framework's effectiveness and its alignment with the core research objectives.

1) Rationale for effectiveness - a new paradigm for human-AI collaboration: The SUNDUS framework presents a new paradigm for human-AI collaboration, as its design directly challenges the assumption supported by the OMAR study: that human operator performance must be optimized primarily through training. Our novelty lies in proposing an alternative, scalable path to high performance through systematic design optimization. The proposed features of SUNDUS are not mere enhancements; rather, they are thoughtfully designed options based on scientific foundations aimed at enhancing human cognition in ways previously only achieved through extensive training. SUNDUS challenges an existing assumption and provides a practical implication for designers and organizations. Furthermore, it will also significantly improve decision-making, reduce cognitive load, and enhance human factors in the collaborative environment between operators and AI."

2) SUNDUS challenges an existing assumption: The SUNDUS framework challenges a prevalent assumption, supported by the OMAR study, that human operator performance must be optimized primarily through training to achieve effective human-AI collaboration. However, SUNDUS proposes a different approach to achieve the same result through system optimization (through intuitive and transparent design). This alternative path to this prevalent assumption is likely more scalable. Shifting the burden of adaptation from the user to the technology has significant implications for how AI tools are designed and deployed, especially in environments where extensive training is unavailable.

3) Practical implications for designers and organizations: Given the SUNDUS model, those working in AI technology design can derive more efficient designs. For example, the SUNDUS approach calls for moving beyond simple, single-sensory user interfaces to information-rich, multisensory designs that support more efficient communication and interaction with the user. For organizations, the SUNDUS model offers opportunities to reduce the costs and time associated with specialized training, enabling faster and more widespread adoption of complex AI systems.

4) Enhanced decision-making: By providing context-sensitive guidance and natural-language explanations, SUNDUS provides an untrained operator with the rationale to make informed decisions. SUNDUS alerts are based on evidence-based arguments (such as visual attributes and confidence scores), enabling an untrained operator to make an accurate and confident decision like a trained operator. The lack of a pre-existing mental model is a defining characteristic of SUNDUS's decision-making process, going beyond automating simple tasks and facilitating true collaboration by making AI logic transparent and actionable.

5) Reduced cognitive load: According to the principles of the AMID framework, underutilized sensory channels (such as auditory and tactile cues) are utilized. Applying this principle to the SUNDUS system reduces the burden of continuous visual monitoring by freeing up the operator's cognitive resources. When the system triggers a critical anomaly alert, the operator's attention is immediately drawn to the system's multisensory alerts, helping to prevent the cognitive overload and fatigue common in sustained monitoring tasks. Furthermore, the system performs the intensive anomaly detection task, allowing the operator to focus their limited cognitive capacity on the more important task of making a final judgment and taking appropriate action.

6) Enhanced "human factors" aspects: Fundamental human factors are influenced by the SUNDUS framework's focus on transparency and intuitive design. For example, using visual attributes, such as color, saturation, and transparency, to convey trust and importance (according to Pesantz et al.'s paper)

supports understanding system outputs. Similarly, adopting an intuitive communication style using simple, elegant language is a solid foundation for building appropriate trust with AI. The trust built here is not blind, as it is calibrated and based on the system's verifiable explanations. Empowering the operator with a system that is easy to understand enhances trust and increases satisfaction, transforming AI from a fearsome black box into a collaborative and trustworthy partner.

7) Increasing transparency: Transparency is a key factor influencing trust, and a fundamental principle of the SUNDUS framework. Transparency in SUNDUS is achieved by replacing a simple alert (a standardized short message) with an informative explanation written in natural operator language and clearly and intuitively presenting the necessary information. This design makes the logic of the SUNDUS AI system completely transparent. Because transparency is a distinct and inherent addition compared to the OMAR system, it will be tested as a key independent variable to determine its correlation and impact on trust and system efficiency.

8) Enhancing overall performance: As with any system, the features and design principles of the SUNDUS framework combine to improve the overall performance of the human-AI team. This goal is achieved by focusing on transparency and intuitive design and ensuring the team can be empowered to improve its performance even with limited or no training. Fulfilling this promise will significantly contribute to the field of human-AI interaction.

9) Between a potential limitation and a counterargument: Despite the promising features of the SUNDUS framework, it may unintentionally lead to misalignment of trust. Could a sound system like SUNDUS lead operators to become complacent and unquestioningly trust the system's "practical recommendations"? SUNDUS can mitigate this risk because its core principle is transparency. Furthermore, its recommendations have a solid logical basis (e.g., natural language interpretations and reliance on established procedural processes). Thus, the system enables operators to understand, analyze, and then make a critical and informed final judgment, rather than mindlessly obey.

V. Conclusion and Future Work

The SUNDUS framework is a conceptual blueprint for a new approach to human-AI collaboration, grounded in the fundamental idea that intuitive system design can serve as a substitute for a lack of formal operator training. While OMAR demonstrated that training is essential, this work argues that the burden of adaptation can be shifted to the system. To help bridge this gap, we strategically leverage two papers on the AMID and principles and the "Visual Representations of Meta-Information" to present the SUNDUS framework, which offers a human-centric solution. The proposed framework aims to achieve similar performance and confidence levels without extensive training. It proposes a system that effectively enhances human perception and understanding through multisensory cues and transparent, informative visual features. In the SUNDUS framework, unintelligible alerts are replaced with a comprehensive, context-sensitive guidance system, reducing cognitive load, improving decision-making, and enhancing and calibrating confidence appropriately. Ultimately, the SUNDUS framework is designed to shift the burden of adaptation from the user to the system, thereby expanding access to advanced AI, increasingly useful in high-risk environments.

A. Future Work

As a conceptual model, confirming the effectiveness of the proposed SUNDUS framework for enhancing performance in a human-AI collaboration environment requires rigorous experimental validation. Therefore, a critical next step is to conduct a practical experiment to test the hypotheses outlined in this paper. This experiment will compare SUNDUS (a transparent state of AI with rich features) and OMAR (a black box).

Furthermore, future research directions could build on this foundational work in several ways, including advanced multisensory exploration: exploring the specific effects of different multisensory cues (e.g., different auditory tones, distinct tactile patterns) on trust and performance could add valuable insights to improve the current framework. Longitudinal studies are also needed to explore how trust in SUNDUS evolves. Such a longitudinal study could determine whether the initial intuitive trust generated by the system is sustainable or whether operators will revert to traditional training. We also recommend conducting studies to customize the current proposed system. Research into adapting the system could potentially explore the flexibility of the SUNDUS framework by allowing operators to specify their preferences for increased transparency, choose the type of multisensory alerts, or even adjust the level of detail provided in explanations using language that matches their needs and business context. Such studies can uncover the link between personal augmentation and human-AI collaboration.

References

[1] B. Shneiderman, "Human-centered artificial intelligence: Reliable, safe & trustworthy," *International Journal of Automation and Smart Technology*, vol. 10, no. 4, pp. 183–195, 2020.

[2] A. Nair and et al., "The ai index report 2024: Measuring trends in artificial intelligence," 2024, accessed 2024. [Online]. Available: https://aiindex.stanford.edu/report/

[3] M. Ameen, "A novel approach to a hybrid security system using operator machine augmentation resource (omar)," Doctoral dissertation, Iowa State University, 2024.

[4] Partnership on AI, "Human-ai collaboration framework & case studies," Website, 2019, accessed 2019-09-25. [Online]. Available: https://partnershiponai.org/paper/human-ai-collaboration-framework-case-studies/

[5] R. T. Stone, A. M. Bisantz, J. Llinas, and V. Paquet, "Augmented multisensory interface design (amid): A human-centric approach to unisensory and multisensory augmented reality design," *Journal of Cognitive Engineering and Decision Making*, vol. 3, no. 4, pp. 362–388, 2009.

[6] A. M. Bisantz, R. T. Stone, J. Pfautz, A. Fouse, M. Farry, E. Roth, A. L. Nagy, and G. Thomas, "Visual representations of meta-information," *Journal of Cognitive Engineering and Decision Making*, vol. 3, no. 1, pp. 67–91, 2009.

[7] F. Chen, Q. Chen, X. Ding, and Z. Li, "Deep Neural Network Uncertainty: A Comparative Study of Quantification Methods," *Journal of Machine Learning Research*, vol. 24, pp. 1–41, 2023.

[8] Y. Lu, L. Wang, L. Jiang, Z. Xu, and H. Li, "The Role of Trust and Transparency in Human-AI Collaborative Decision-Making," *Frontiers in Psychology*, vol. 12, p. 706530, 2021.

[9] A. Holzinger, G. Langs, M. Kenk, P. Kieseberg, and A. Tjoa, "Explainable AI (XAI) for decision support in high-risk domains," *Machine Learning and Knowledge Extraction*, vol. 5, pp. 906–934, 2023.

[10] J. D. Lee and K.-G. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004.

[11] S. Amershi, D. Finai, L. Clark, J. Davis, C. Hendershott, H. Hochheiser, W. Lasecki, B. Lee, X. Li, J. Lin, J. Manes, R. Mahajan, M. R. Morris, I. Naim, J. R. Smith, and S. Studer, "Guidelines for Human-AI Interaction," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*, 2020, pp. 1–20.

[12] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, J. Gil-López, D. Molina, V. Benjamins, F. Herrera, and J. A. Martínez-López, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges," *Information Fusion*, vol. 58, pp. 82–115, 2020.

[13] J. Tesch, R. Tesch, D. Seeliger, F. Spelz, and M. Schöbel, "Deep Learning for Multi-Modal Human Activity Recognition: A Survey," *IEEE Access*, vol. 11, pp. 28 003–28 020, 2023.

[14] D. Gunning, E. Vorm, S. Piatnitski, M. Turek, and K. Al-Khateeb, "XAI: Explanation Strategies for Artificial Intelligence," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, pp. 4789–4801, 2021.

[15] F. Doshi-Velez and B. Kim, "Towards a unified evaluation of interpretability in machine learning," *arXiv preprint arXiv:1709.06830*, 2021.

[16] Y. Rastegar, B. Mirmahboub, M. Saadat, M. Ahmadi, and H. Sadr, "Explaining Deep Learning for Crowd Behavior Analysis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2021, pp. 3103–3112.

[17] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence*, vol. 304, p. 103639, 2022.

[18] Y. Zhou, C. Zhang, X. Chen, and X. Liu, "Visualizing Deep Learning: From Feature Visualization to Interpretability," *Frontiers in Artificial Intelligence*, vol. 3, p. 55, 2020.

[19] P. Wang, S. Feng, C. Liu, L. Liu, J. Lin, S. Li, S. Guo, and Z. Lv, "A Survey on Haptic Feedback Systems in Augmented Reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, pp. 4455–4467, 2022.

[20] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 30, no. 4, pp. 286–291, 2000.

[21] T. Alafif, M. Jassas, A. E. Abdel-Hakim, G. Alfattni, H. Althobaiti, M. Ikram, A. Alharbi, H. Alsharif, M. AlShamrani, E. Alharbi, T. Alsubait, A. Alhawsawi, B. Alsolami, and K. Khayyat, "Toward an Integrated Intelligent Framework for Crowd Control and Management (IICCM)," *IEEE Access*, vol. 13, pp. 58 559–58 575, 2025.

[22] K. A. Hoff and M. Bashir, "Joint cognitive systems: A relational perspective on human-automation trust," *Frontiers in Psychology*, vol. 12, p. 638202, 2021.

[23] T. A. Slocum, D. C. Cliburn, J. J. Feddema, and J. R. Miller, "Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance," *Cartography and Geographic Information Science*, vol. 30, no. 4, pp. 299–317, 2003.

[24] R. Rosenholtz, Y. Li, and L. Nakano, "Measuring visual clutter," *Journal of Vision*, vol. 7, no. 2, pp. 1–22, 2004.

[25] P. I. Terrence, J. C. Brill, and R. D. Gilson, "Body orientation and the perception of spatial auditory and tactile cues," in *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*. Human Factors and Ergonomics Society, 2005, pp. 1663–1667.