Ajna: A Wearable Shared Perception System for Extreme Sensemaking

MATTHEW WILCHEK, Department of Computer Science, Virginia Tech, USA
KURT LUTHER, Department of Computer Science, Virginia Tech, USA
FERAS A. BATARSEH, Department of Biological Systems Engineering, Virginia Tech, USA

This paper introduces the design and prototype of Ajna, a wearable shared perception system for supporting extreme sensemaking in emergency scenarios. Ajna addresses technical challenges in Augmented Reality (AR) devices, specifically the limitations of depth sensors and cameras. These limitations confine object detection to close proximity and hinder perception beyond immediate surroundings, through obstructions, or across different structural levels, impacting collaborative use. It harnesses the Inertial Measurement Unit (IMU) in AR devices to measure users' relative distances from a set physical point, enabling object detection sharing among multiple users across obstacles like walls and over distances. We tested Ajna's effectiveness in a controlled study with 15 participants simulating emergency situations in a multi-story building. We found that Ajna improved object detection, location awareness, and situational awareness, and reduced search times by 15%. Ajna's performance in simulated environments highlights the potential of artificial intelligence (AI) to enhance sensemaking in critical situations, offering insights for law enforcement, search and rescue, and infrastructure management.

CCS Concepts: • **Computing methodologies** → **Multi-agent systems**; • **Augmented Reality**; • **Wearable Technology**; • **Computer Vision** → Object Detection;

Additional Key Words and Phrases: Object Detection, Augmented Reality, Human-Computer Interaction, Human-in-the-loop, Distributed Systems

1 INTRODUCTION

Sensemaking is a decision-making process that involves creating plausible explanations for human actions [98]. This is particularly important in emergency situations where environmental constraints can hinder life-saving actions [28]. The term *extreme sensemaking* refers to sensemaking performed under extreme conditions, such as navigating a partially collapsed building post-earthquake that requires quick decision-making for locating survivors. In these contexts, decentralized "edge" networks support peer-to-peer communication and efficient parallel searches by distributed teams, eliminating the need for a central authority delegating critical information [25].

Extreme sensemaking becomes more complex in environments without the use of electronic communication channels, such as radio, internet, or cell coverage. There are also complications when an operator needs to react stealthily or uses mobile devices that interfere with operating other handheld equipment such as tools, radios, or firehoses. To address these challenges, our research investigates the development of reliable and effective wearable technology to support sensemaking tasks in domains such as fire and law enforcement (see Fig. 1), maintenance, and search and rescue efforts [31, 58].

Authors' addresses: Matthew Wilchek, mwilchek@vt.edu, Department of Computer Science, Virginia Tech, Arlington, VA, USA; Kurt Luther, Department of Computer Science, Virginia Tech, Arlington, VA, USA, kluther@vt.edu; Feras A. Batarseh, Department of Biological Systems Engineering, Virginia Tech, Arlington, VA, USA, batarseh@vt.edu.

ACM 2160-6463/2024/9-ART
https://doi.org/10.1145/3690829

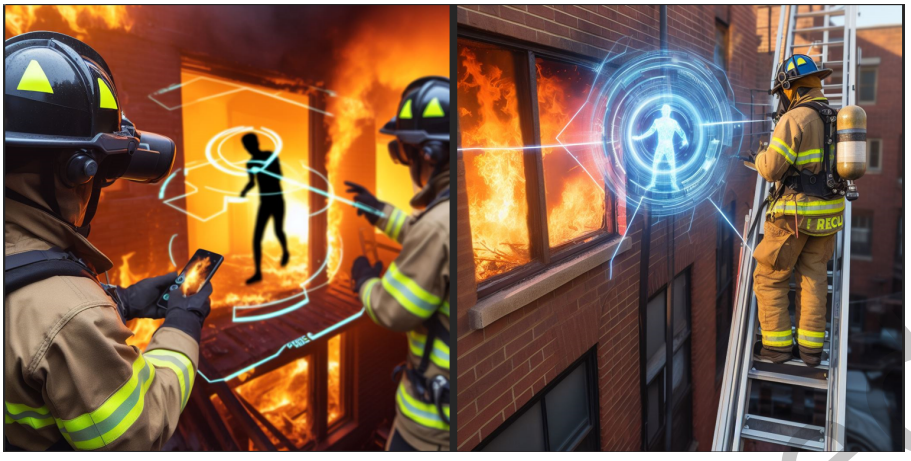


Fig. 1. An illustration of Extreme Sensemaking (generated by DALLE-3). In this example, the two firefighters on the left image detected a real-time, 3D location of a person in a burning building accessible via a ladder for rescue. The location information, generated by one of the firefighters inside the building, is then shared with the ladder crew outside on the right image, allowing them to swiftly comprehend the distressed individual's exact whereabouts.

As wearable Augmented Reality (AR) devices like Apple Vision Pro and Microsoft HoloLens gain popularity, the development and testing of innovative applications for these devices are also on the rise [8]. These AR devices employ a suite of technologies, including photographic and depth cameras, accelerometers, gyroscopes, and magnetometers, to accurately capture the user's surroundings. The combination of these resources with edge computation can facilitate object detection and localization in three-dimensional space, allowing for enhanced situational awareness [1, 9, 51, 93, 97]. In recent years, researchers have begun exploring the integration of embedded Artificial Intelligence (AI) algorithms on AR devices for accurate object-detection tracking and real-time scene understanding [24, 33, 52]. However, a common challenge with these technologies is scaling up AI-driven sensemaking for concurrent users in larger environments due to the relatively low resolution of AR sensors limiting the detection of distant objects [35]. Section 3 further discusses these issues and proposes potential workarounds.

In this paper, we introduce *Ajna*, a novel shared perception system designed to overcome the current limitations of AR devices, enabling scalable object detection sharing across distances and barriers previously untenable. The term *Ajna* is derived from the Hindu *chakra*, "the third eye", which symbolizes a heightened perception of the world beyond what is visible through two eyes [76]. Named after the concept of a higher awareness, Ajna utilizes the Inertial Measurement Unit (IMU) present in most AR devices to measure the relative distance of each user from a mutually agreed physical point, allowing object detection information to be shared and viewed from this common understanding with minimal spatial drift. Our design offers a collective perception capability, effectively serving as a "third eye" for extended situational awareness.

Scope. The purpose of this work is to improve object recognition and AI-driven sensemaking for teams or crowds to address the difficulties associated with extreme sensemaking in the AR domain. Current AR depth sensors and cameras have limitations that prevent users from perceiving objects beyond their immediate surroundings, behind barriers, or on multilevel structures. Due to these constraints, additional methods like QR code markers, specialized antennas, or geo-rectification techniques are commonly required to overcome perceptual obstacles [9]. Our shared perception application, Ajna, is designed for high-intensity, limited-scale

operational zones such as multilevel structures or compact metropolitan blocks. As described in Section 4, Ajna integrates its three basic modules (SpatialSense, CoInferX, and PercepShare) to support collaborative operations.

Our evaluation studies tested and validated the integration of Ajna, creating a shared perception system that operates without the need for extra supportive infrastructure. The Crowds-In-The-Loop method [18] allows users to update detections for moving objects or identify and update the relevance of detected objects, such as stationary injured persons. This detection flexibility and human-centered design enhance accuracy and utility. Although Ajna's precision decreases in expansive areas, our findings demonstrate improved object detection and easier location sharing among teams. A user study with controlled trials assessed Ajna's sensemaking capabilities in a simulated multilevel urban search and rescue operation. The results showed that participants using Ajna required, on average, 15% less time for tasks involving the identification and localization of simulated victims. Participants also rated Ajna's user interface and overall system as acceptable based on the System Usability Scale (SUS). Overall, we found evidence that Ajna can support collective spatial understanding and facilitate extreme sensemaking, with potential applications in real-world SAR operations, law enforcement training, and navigation in complex terrain when used on ruggedized AR devices [84].

Summary of Contributions. Building on previous studies [4, 34, 53], our work makes the following novel contributions: (1) We introduce Ajna, a novel shared perception application for AR devices that can enhance team sensemaking, particularly when traditional communication channels are limited. (2) Ajna demonstrates how shared spatial references may be used in AR efficiently without QR codes or additional edge hardware and cloud services. With this design approach, occluded objects are tracked smoothly, and situational awareness is enhanced in unknown environments, improving system feasibility and accessibility. (3) We evaluate Ajna in a simulated search and rescue environment, demonstrating a significant decrease in task completion times. Using empirical evidence from our study, we illustrate Ajna's utility in developing collective decision-making in real-world environments, emphasizing its practicality and efficacy.

Although our design for this study is based on the Microsoft HoloLens 2, it is not limited to this device, allowing for broader applicability. Ajna overcomes the constraints of unreliable internet and eliminates the need for a pre-established infrastructure or off-device capabilities, making it readily deployable to AR devices [61].

2 RELATED WORK

Augmented reality devices offer transformative advantages over traditional devices, primarily through their ability to superimpose digital information such as images, data, and animations onto the physical world, thereby enhancing a user's real-world experience. This technology is especially advantageous in fields requiring complex interaction with the environment or data analysis, such as medicine, engineering, education, and military operations [27, 36, 42, 57]. Traditional form factors, like desktop and laptop computers and handheld devices, often constrain users to screens, limiting the interaction to 2D information presentation and demanding divided attention between the task and the device [99]. In contrast, AR devices integrate information directly into the user's environment, allowing for more natural interaction and multi-sensory engagement, which is crucial for tasks requiring spatial understanding or real-time decision-making [5, 8, 11]. Furthermore, recent studies indicate that AR can reduce cognitive load and human error as users navigate and interpret information more intuitively within their field of view, rather than diverting attention between the real world and a screen [91]. These advancements underscore AR's potential to significantly enhance performance across various disciplines, surpassing the capabilities of traditional devices.

Building on this momentum, AR headsets are increasingly being utilized in search and rescue operations and emergency services due to their unique ability to enhance situational awareness and operational efficiency in high-stakes environments. These devices superimpose crucial data, including maps, victim locations, and hazardous area markers, directly onto the rescuers' field of vision. Furthermore, Ong et al. showcased AR's role

in expediting victim identification and improving navigation under difficult conditions [79]. Similarly, Luskas et al. developed a HoloLens application tailored for SAR based on SAR professionals' input. The application allows users to annotate surroundings, mark points of interest, and access vital SAR information, with plans to incorporate more collaborative features for shared holographic data [58]. Chalimas et al. advanced this technology further by integrating an infrared camera with the HoloLens, testing its application in fire rescue scenarios [12]. These innovations hold the potential to enhance victim detection in multi-story buildings, allowing for rapid information sharing among rescuers, which is crucial in time-sensitive rescue missions.

The effectiveness of these AR applications in real-world SAR scenarios is also contingent on the compatibility of the technology with the specialized gear used by SAR professionals. The development of ruggedized HoloLens versions, such as the firefighter-specific C-THRU helmet by Qwake Technologies, the U.S. Army's Integrated Visual Augmentation System (IVAS), and the Military Augmented Reality System (MARS) employed by China's People's Liberation Army could pave the way for more comprehensive field studies and the broader implementation of AR in SAR operations [47, 75, 84]. These advancements demonstrate the growing feasibility and importance of AR technology in enhancing the capabilities of SAR professionals.

The advancements in ruggedized AR tools, essential for SAR operations, lay the groundwork for another crucial aspect of AR applications in emergency scenarios: object detection and collaborative data sharing. Object detection using state-of-the-art AR systems is not a new idea. Works such as [13, 15, 97] investigate the use of advanced machine learning models on the HoloLens 1 or 2 specifically for object detection. These studies even incorporate depth perception capabilities of the HoloLens or other AR devices for object location or use networked servers to aid in the computationally intensive object detection task. However, these works implement object detections solely on a single user (a single HoloLens) or between a single user and an edge server to distribute computation load. These detections' accuracy completely relies on that single user's field of view (FOV). Research has also shown the feasibility of using object detection on AR devices for rapid damage detection, crack inspection, structural health monitoring (SHM), excavation, and indicating structural displacements [21, 45, 46]. However, if a repair crew could communicate and share their detections with each other as a collaborative application, then it would be possible to validate the detections between what each other sees in real-time.

Another area that has seen research in integrating AI with AR is 3D scene reconstruction. 3D scene reconstruction is the process of generating a digital 3D model of a real-world scene from multiple images or sensor data [81]. The resulting 3D model can be used for various applications, such as virtual reality, augmented reality, and robotics. Tateno et al. demonstrate a novel monocular SLAM system that uses a convolutional neural network (CNN) for learned depth prediction to generate 3D reconstructions of scenes in real-time [89]. The CNN estimates the depth of each pixel in the input image and generates a dense depth map for the reconstruction. However, the authors only propose the work can be used for augmented reality devices and only evaluate the work in controlled experiments on a standard computer desktop CPU.

Similarly, Wald et al. propose a real-time fully incremental scene understanding system originally designed for mobile systems [96]. The proposed system was designed to learn and adapt to new objects and environments incrementally; however, it was only evaluated on a single Google Tango device, which is an AR device that requires hands-on control. A common theme around these two notable works and others [e.g. 14, 20, 74, 86, 90] is the proposal that these novel contributions in 3D scene reconstruction can aid in sensemaking for real-world scenarios with AR, but fail to include any kind of evaluations that back-up these claims.

As mentioned briefly in Section 1, the existing body of literature on the integration of AI with AR primarily focuses on individual device usage [95]. However, recent research has witnessed a surge in investigating the application of human-in-the-loop (HITL) learning for object detection on AR devices, aiming to enhance the process of sensemaking. The involvement of human users through HITL learning has proven to enhance object detection performance [101]. Nevertheless, the majority of these studies have concentrated on single-user scenarios. For example, Hoppenstedt et al. utilized the HoloLens and voice commands to label objects for training

convolutional neural networks (CNNs), employing an active learning component [38]. However, this HITL approach is designed for a single user and does not support collaborative, distributed multi-user environments. Additionally, there is a need to further validate the assurance of their model. Furthermore, the computation for object detection occurs remotely on a web server rather than on the physical HoloLens device or an alternative edge device. This raises concerns regarding operating in real-time environments with poor or insecure signals. Several authors have proposed similar concepts on different platforms, including mobile devices, web applications utilizing crowdsourced intelligence, and electrocardiogram headsets (e.g., [40, 51, 55, 56, 70, 80, 82, 102]). Abraham et al. introduced a HITL pipeline to enhance object detection using drones and improve overall detection reliability [2]. Their autonomous system is designed to adjust the level of human involvement based on perceived limitations in the drone's detection capabilities. However, their proposed technique lacks integration with a connected network of detections for a shared perception system among multiple drones.

In 2022, Smith et al. developed a software application for the HoloLens 2 to assist in bridge inspections, which primarily includes annotation and spatial markers for areas that need repair [85]. However, the application is primarily driven by human input. Future research is discussed in the next steps for the need to incorporate corrosion state and crack detection algorithms for automated annotation. How objects are visualized on these systems can also impact a user's trust in the system. For example, Guan et al. developed an object detection application for the HoloLens that visualizes detections with 3D bounding boxes but was only able to achieve 30 frames per second (FPS) [34]. Their work extends some of the advancements in incorporating depth data from AR headsets for object detection [3, 50]. However, all of these concepts were not HITL-focused or capable of scaling to support multiple, concurrent users.

In addition, these prior works do not conduct evaluations to measure users' trust or interpretability of the systems. Wickramasinghe et al. discuss when users can see the output and outcomes of a machine-learning model through visualizations, it can enhance their understanding of how the model works and the quality of its predictions [100]. For example, transparent visualizations enable users to assess whether the model's predictions align with their expectations or match their domain knowledge. If users can interact with visual representations of the data and explore different aspects of the model's outputs, it enhances their involvement and confidence in the results. Engaging visualizations can create a sense of ownership and collaboration between the user and the AI model [37, 59, 87]. Therefore, we sought to create transparent and interactive visualizations in the development of our prototype.

3 CHALLENGES AND KEY INSIGHTS

In this section, we explore the design challenges associated with developing a system that facilitates extreme sensemaking, along with the essential insights that enable Ajna to effectively support collaborative object detection, a fundamental capability for such a system.

3.1 Design Challenges

An initial attempt at designing a system to share detections might focus on merely packaging the metadata of a detection generated locally by a device, such as the location, label, and detection confidence, and sending this data on a network to others. This technique, however, would suffer from the inherent challenges described below.

Challenge 1: Large-scale AR spatial reference is non-trivial. AR devices are designed to understand and function in relatively small spaces, such as rooms and hallways. Their onboard Simultaneous Localization and Mapping (SLAM) techniques are designed to isolate identifiable features of these spaces and to use them as static reference points to orient the location of the device, and that of any holographic material present in this environment [13]. This understanding can even be translated into "spatial anchors," which can be passed among similar devices to create a unified frame of reference to allow for collaborative AR applications [17]. However,

when the device is presented with larger spaces that exceed the capability of its depth sensors and cameras to map a given space, the device must then rely on other means to orient itself and share this orientation with others [54]. Some solutions to this problem require special modifications to the environment (e.g. QR codes) or rely on matching outdoor features to overhead satellite imagery [69]. This assumes that either the user is outdoors and/or in an environment with these special physical modifications. Neither of these can be assumed if the user moves between indoor/outdoor environments or is in a new area without pre-positioned modifications. This problem prevents one or more users from establishing a common understanding of a large-scale space and inhibits collaborative activities in these areas.

Challenge 2: Object Detection for AR is inherently limited in scope and distance. Object detection on AR headsets has the potential to revolutionize the field of enhanced sensemaking by providing real-time, interactive, and immersive experiences that can help users understand complex data and information. However, key technical challenges continue to exist preventing a refined capability. One such challenge is a limited view range. A limited FOV can make it difficult to detect objects outside the user's line of sight. This physical sensor limitation of the device makes object detection models struggle to resolve inference on small or distant objects due to limited resolution [32]. As a result, object detection capabilities on most cameras including for AR devices, lead to missed detections and incomplete scene understanding [22, 39]. A wider FOV can make it more difficult for users to focus on specific virtual objects or naturally interact with them. AR headsets can also face challenges in object detection due to occlusions and object interactions, such as objects appearing behind walls or interacting with each other. This can make it difficult to accurately detect objects and understand the scene [23, 33]. Ultimately, the combination of these two challenges prevents concurrent users from using efficient object detection collaboratively in AR.

3.2 Key Insights

Modern AR devices are designed to navigate their environments using a combination of SLAM and information from an onboard suite of sensors that record the device's motion and rotation among other things, called an Inertial Measurement Unit (IMU) [83]. This IMU records when the user moves forward, laterally, or vertically (e.g. by using stairs), and this information is used to determine the user's new position relative to the device's understanding of its physical environment. If two or more devices share a common understanding of a single point in a mutually understood physical space (e.g. a single room), we can exploit this common reference by using the IMU data to measure the relative distance from this waypoint as the user moves away from it. This suggests that *we can overcome the large-scale environment problem by allowing each user to move from this common spot, measuring the distance they have moved in 3D, and permitting collaborative, accurate, and large-scale collaboration in AR without QR codes or referencing satellite imagery addressing Challenge 1.*

The computational limitations of AR headsets were observed in the Guan et al. study, which is made worse by the multitude of sensors needed to support AR features [34]. The low quality of depth images obtained from AR headset device cameras is a clear indication of this limitation, which reduces the effectiveness of 3D object detection algorithms [94]. It has been demonstrated that HITL learning greatly improves system performance by having a human expert help the system adapt in real-time to new situations [101]. In particular, Razeghi showed that applying HITL learning could improve performance in object detection systems in variable environments by up to 25% [77]. Our crowd-in-the-loop method encompasses HITL learning within collaborative AR spatial referencing, presenting a potential solution to key challenges in collaborative object detection within AR contexts [18]. The use of Ajna by multiple users simultaneously for spatial referencing widens the FOV for object detection. As more users connect through Ajna, it should theoretically become possible to detect and share objects over a larger area [26, 103].

Furthermore, user participation makes accurate localization and timing for object detection inferences possible. This capability simplifies occluded object location identification from the spatial referencing associated with detections, thereby facilitating rapid scene understanding. Rapid scene understanding can be especially helpful for detecting where objects of interest may be despite not being in the immediate area, like those that are hidden by structural barriers, from a distance through several walls, or on different floors of buildings. *With this insight, we can overcome the limitations proposed in Challenge 2.*

4 SYSTEM DESIGN

In this section, we propose a design that addresses the challenges in Section 3.1 by exploiting the key insights in Section 3.2. We first break down our design goals for the system, and core components that were brainstormed for the concept, then detail our initial approach before implementation similar to the guidelines presented in the Co-Creative Framework for Interaction design (COFI) [78].

4.1 Design Goals

After analyzing related research, summarizing challenges, and evaluating key insights in Sections 2 and 3, we created the following design goals to explore the ability of AR devices for extreme sensemaking.

(1) **Scalability:** Multiple, concurrent AR devices must form a peer-to-peer (P2P) network to communicate spatial maps (i.e., SLAM data), holograms, and object detection results. The AR devices and network should be capable of supporting up to 3 to 5 devices at one time; theoretically more.
(2) **On-Device Inference:** Each AR device is capable of efficiently processing a frame from its onboard camera, passing it to a pre-trained object detection model, and post-processing the results of any detections that exist.
(3) **Usability:** The system must provide a user interface (UI) that gives each AR user control in guiding when and where to execute the object detection software, then how to communicate results to other concurrent AR users part of the P2P network. This should be an intuitive HITL workflow that enhances collaborative sensemaking and minimizes learning curves.
(4) **Trustworthy:** The system should have reliable and optimized object detection models trained on objects of interest per application context. The models should be capable of processing frames accurately with a low false-positive rate to provide trustworthy insights that aid in sensemaking.
(5) **Modularity:** Each AR device should have the option to swap out different object detection models with others. Users should have the option to select different models in the UI. Regardless of what model is used on the device, Ajna should still function as designed. This design goal will support multi-modal inference for the system.
(6) **Minimal Hardware:** Other than a local network connection and the AR devices, no additional hardware should be used to avoid the need for additional services or infrastructure. Also allows the system to be agile and adaptable to extreme environments.

We require a design that can locate objects and share representations of them in 3D space at distances beyond the same local area. Additionally, we need a system that can integrate the HITL approaches needed to add human cognition and decision-making to this system by allowing human choice and intervention. Each capability will be essential to the other (i.e., non-discrete; one cannot exist without the other). After planning our design goals, we translated our requirements into three core components that will be critical in our implementation. We summarize these components as follows:

(1) **SpatialSense:** We capitalize on the use of the device's onboard Inertial Measurement Unit to track the user's movement relative to a shared, common physical location. We use this location to infer the location

of any detected object relative to the user. This technique extends the device's intended functionality and allows for long-range indoor/outdoor tracking of detected objects by multiple users.

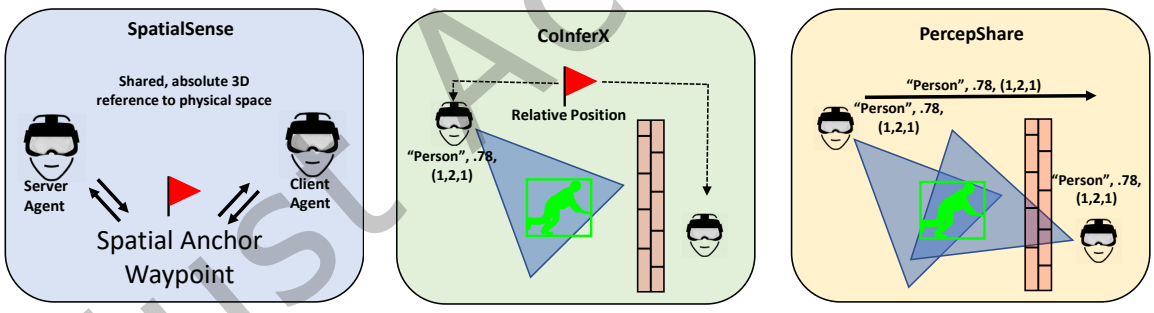
(2) **CoInferX (Collaborative Inference Experience):** Multiple, concurrent users of AR headsets each using Ajna to decide where to look and execute AI inference. The object's location is spatially recorded with its corresponding detection results by giving users control over when and where to run object detection algorithms. *This component is dependent on component 1.*

(3) **PercepShare (Perception Sharing Module):** Each Ajna user is part of an integrated system. When one user shares an object's physical location and corresponding detection details for something of interest, all users of the system can visually see the information. This component allows users to see far-away objects through physical walls and ceilings that would otherwise be undetectable in a large-scale environment, along with the most accurate detection details inferred by other users. In the case of two or more users sharing the same location, the system functions in the same way. This could be useful in environments where, even when co-located, users may not be aware of the entirety of their surroundings. *This component is dependent on components 1 and 2.*

We deliberately designed Ajna with a focus on software capabilities to ensure its applicability in challenging environments if deployed to ruggedized AR devices. A key emphasis within Ajna lies in its AI capabilities. We present greater detail in the design, implementation, and evaluation studies for each of these components in Sections 5, 6, and 7.

4.2 Our Approach

We leverage the inherent functionality of AR devices in a way that extends the normal functionality of the device and facilitates extreme sensemaking. We do not claim that any of these techniques are in and of themselves a new invention, but a re-purposing of existing technology in a way that uniquely meets a need that has yet to be realized.



(a) Step 1. (b) Step 2. (c) Step 3.

Fig. 2. This figure outlines the proposed software design required to address the challenges posed in Section 3. In Step 1(a), Ajna users exchange an absolute spatial reference to a common physical point we refer to as a "waypoint". After recording each user's relative distance from the waypoint, in Step 2(b), one Ajna user detects an object and selects it for distribution to other users. In Step 3(c), users exchange information about the location of the detection using their relative distance from the waypoint.

SpatialSense. To exchange the singular, absolute reference to physical space, we leverage Spatial Anchors, an abstraction made possible by the ability of AR devices to sense and map a physical space [64]. Each device creates, maintains, and exchanges these pieces of information differently, but the end product of this exchange must be a

coherent understanding of the physical environment on two or more devices. We use this to establish the common "waypoint" as shown in Fig 2a. This waypoint must be generated from a common point, that both devices have observed and mapped. This requires both devices to be in the same proximity. However, this common frame of reference is not in and of itself sufficient for solving *Challenge 1*, and we require more information about the movement of the user to locate them relative to this anchor.

Even with the waypoint, we do not yet know where each user is located in relation, especially when the users move away from the waypoint and into areas obscured from each others' view. We now leverage the capability of the IMU on the device to measure the relative distance each user travels after exchanging the waypoint in the same physical space. This exploitation of the inherent functionality of the device is how we can exchange understanding of each user's location, and set the conditions for extreme sensemaking. The combination of the waypoint and this relative understanding of a user's location solves *Challenge 1*.

CoInferX. Next as in Step 2 in Fig. 2b, we must detect objects of significance. While the detection of objects in images has been studied extensively, as demonstrated in works such as [104], the challenge in AR-specific research lies in transforming the two-dimensional object detection results obtained from these techniques into three-dimensional space. To achieve this, AR devices offer a variety of tools. By utilizing metadata from each captured image and the two-dimensional location of the detected objects within the image, it is possible to project these locations onto a three-dimensional space, as demonstrated in previous AR-specific research such as [13, 16]. Each captured frame contains meta-data, known as CameraIntrinsics [62], to take the two-dimensional detection to three dimensions. Similar to how spatial anchors are transferred between the host to client headsets, detection result holograms are also transferred. After a detection is made and a 3-dimensional bounding box hologram is displayed, a user can interact with that bounding box and then manually send it to other system users by tapping it. Additionally, using HITL, we can determine which detected objects are important enough to justify sharing with other users, increasing accuracy and limiting distraction. This technique is designed for stationary objects like injured people. However, the user of the device can update the detection to account for the motion of the object by re-detecting the object at a time of their choosing. Here, we continue to leverage the HITL techniques to allow humans the ability to decide when and where to update these detections.

PercepShare. Finally, as in Step 3 in 2c, we must exchange the detections, including the relative position that is needed to justify the location in areas the other users may not have explored. Design patterns such as cloud environments, near-field communications channels like Bluetooth, or services like Wi-Fi are potential options. However, for our specific use cases outlined in Section 1, it is important to select a communication channel with a sufficient range to communicate around buildings or small outdoor areas without relying on internet access. A local area network appears to be the most suitable choice. Wi-Fi offers the necessary range for small outdoor areas or single buildings, and its infrastructure (i.e., Wi-Fi router) is portable enough to be carried or deployed quickly by device users. The ability to share detections across large distances and to collaborate to realize extreme sensemaking solves *Challenge 2*.

Taking these design considerations into account, we propose a system that leverages the unique capabilities of AR devices to address the challenges presented in Section 3. This system will utilize AR's ability to comprehend physical space, coupled with state-of-the-art object detection models, and supported by local area network systems.

5 AJNA IMPLEMENTATION

This section provides a comprehensive overview of the implementation details of the software that we developed to address our system design goals. Additionally, we highlight how our framework surpasses previous solutions that do not integrate human feedback. In Section 6, we elaborate on our experimental setup that demonstrates the enhanced object detection accuracy and the novel capabilities discovered through our shared perception system.

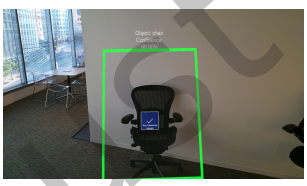
Spatial Anchor Distribution. Spatial Anchors are tools used to create a joint, absolute understanding of a 3D space between two or more devices[64]. In short, these anchors allow two or more users to see the same synthetic object in the same place, regardless of differences in their individual device's understanding of their location. The first step in our design was to generate, transfer, and incorporate these anchors.

While cloud-based tools such as Microsoft's Azure Spatial Anchors (ASA) [17] exist to ease the transfer of spatial anchors, these require a reliable internet connection. The motivating examples discussed in our introduction to extreme sensemaking, Section 1, presented a scenario where an active internet connection may not exist. Even if this assumption was inaccurate, using ASA requires a paid subscription to Microsoft's Azure cloud for each device. We believe not assuming a live internet connection makes the prototype more flexible, as requiring only a working home router is less of a requirement than expecting the device to have an active internet gateway. Keeping this in mind, we choose to implement the spatial anchor transfer (and the subsequent detection transfers) using only local data and connections.

We designed a server and client relationship for connections through simple socket programming. One headset acts as the server with the associated IP address and listens for connection client requests from a local network connection shared by all headsets. Once the host headset receives a request for connection from a client, the host generates a spatial anchor using OpenXR's XRAnchorTransferBatch class [67]. The host then sends this anchor to the client over the socket connection. By replacing a client's spatial understanding, we can guarantee that all physical objects perceived by multiple headsets of the network are anchored to the same physical location.

Object Detection and Transformation. Once a peer-to-peer network has been established and spatial anchors are transferred between devices, we must provide a method for the device to create 3D object detections. Users are provided a choice of interactable buttons to select what object detection model they would run, either YOLOv2, Tiny-YOLOv2, Tiny-YOLOv3, or YOLOv4. Then, each headset will load that model for their detection algorithm, and users can begin detecting objects of interest they see automatically or manually. Ajna can be configured in the developer settings to run the object detection model constantly if desired. However, the default setting allows a user to indicate when to trigger the algorithm to detect by pressing a button in the user interface.

Detection Distribution. This final step involves communicating the detection or detections produced on one device to another. This is done using Universal Windows Platform (UWP) [66] and its socket APIs. Each user is given the freedom to choose which detection they believe is important enough to be distributed. They exercise



(a) User 1 Detects the Left Chair at 90% Confidence.

(b) User 1 Sends the Detection to Agent 2.

(c) User 2 Detects the Right Chair at 77% Confidence.

(d) User 2 Receives the Left Chair Detection from User 1.

Fig. 3. Spatial Anchor Transfer of Object Detections

this choice by selecting the detection using a point-and-click method called an "Air Tap" [63]. Once chosen, the detection is then placed in a queue for transmission to all connected, concurrent users.

In Figure 3, we illustrate this transfer. We can see User 1 detects the left chair with a confidence of 90%, shown with a three-dimensional bounding box. User 2 detects the right chair with a confidence of 77% and is also shown with a three-dimensional bounding box. User 1 then sends the detection of the left chair to user 2, which is received and shown with a blue bounding box. Both users in this test were using the Tiny-YOLOv2 model. Deciding what object detection model to include in the prototype was one of the biggest challenges we observed when we realized implementation would differ for each model on the HoloLens 2.

6 AJNA EXPERIMENTS AND RESULTS

Following the development of Ajna, we conducted multiple experiments to assess its potential in extreme sensemaking with multiple concurrent HoloLens 2 users. Our primary objective was to investigate the interplay of AI within a shared perception system, especially when utilized by multiple concurrent HoloLens 2 users. We carefully selected these experiments to understand AI's opportunities and limitations in scenarios like SAR teams or other collaborative groups engaged in sensemaking. Section 7, validates the effectiveness of Ajna through a usability study from the perspective of a mock search and rescue scenario.

6.1 Software and Hardware Setup

We built the shared perception prototype in Unity 2020.3.16f1, using MRTK version 2.7.2 [65]. The software was deployed on several Microsoft HoloLens 2 HMDs running Windows Holographic for Business Build 20348.1528. All source code was completed in C#.

6.2 Aided YOLO Object Detection (Experiment 1)

In our first experiment, we aimed to evaluate the SpatialSense component of Ajna from our original design goals in Section 4. SpatialSense was designed to improve the confidence and distance of objects detected from an object detection model. To accomplish this, we conducted a coordinated detection test in a small office room using three Microsoft HoloLens-wearing users (Agents) who executed the same object detection model (Tiny-YOLOv2) from the prototype. The users circled a chair located at the center of the room, with User 1 standing 5 yards away, User 2 standing 10 yards away, and User 3 standing 15 yards away from the object, as shown in Figure 4. Our objective was to observe whether the farthest agent who failed to detect the object (either User 2 or 3) would receive the best detection from either User 1 or 2. We recorded the confidence levels observed by each user every 60 degrees and presented the experiment results in Table 1.

User 1's confidence levels varied with different angles despite standing at the same distance from the object. However, they experienced the highest confidence level in detecting the chair compared to Users 2 and 3. The majority of observations by Users 2 and 3 failed to detect the chair initially. Despite this, User 1 chose to share their detection since they had the highest confidence with Users 2 and 3 regardless of their distance and angle. Our null hypothesis was therefore accepted, and we conclude that users of the shared perception system can achieve the most accurate object detection possible. We observed through the first evaluation of SpatialSense that an increase in the number of users of Ajna can ensure that far-away users can still detect objects not originally seen and with higher confidence. These phenomena may begin to address the original distance and accuracy limitations for object detection cited in Section 2. As an initial response to our original hypothesis, extreme sensemaking could be enhanced by a HITL component.

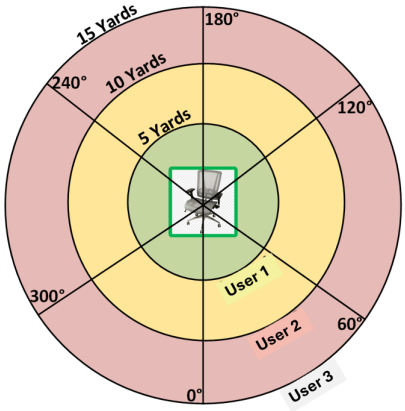


Fig. 4. Distance and Angle Detection Evaluation

Table 1. Shared Perception Detections

User	Yards from Object	Angle of Target	Original Detected Confidence	Observed Confidence from User	Received from User
1	5	0°	70%	-	No
1	5	60°	46%	-	No
1	5	120°	56%	-	No
1	5	180°	43%	-	No
1	5	240°	65%	-	No
1	5	300°	56%	-	No
2	10	0°	0%	70%	Yes
2	10	60°	0%	70%	Yes
2	10	120°	0%	70%	Yes
2	10	180°	0%	70%	Yes
2	10	240°	0%	70%	Yes
2	10	300°	0%	70%	Yes
3	15	0°	0%	70%	Yes
3	15	60°	0%	70%	Yes
3	15	120°	0%	70%	Yes
3	15	180°	0%	70%	Yes
3	15	240°	0%	70%	Yes
3	15	300°	0%	70%	Yes

6.3 Multi-Modal Inference for Object Detection (Experiment 2)

Our second experiment aimed to investigate the potential of Ajna for enhancing context awareness through the detection of different objects by multiple concurrent users. This experiment was designed to test the modularity goal of the system for our three components. We conducted the experiment in a residential setting, where three users wearing Microsoft HoloLens 2 detected objects using different object detection models. Users 1 and 2 used Tiny-YOLOv2, while User 3 used YOLOv4, which is capable of detecting up to 80 different objects compared to

Tiny-YOLOv2, which is limited to detecting up to 20 objects. Each user began in the center of a room, with each facing a different area that did not overlap with the others. Specifically, User 1 focused on the living room, User 2 on the dining room, and User 3 on the kitchen. Upon detecting objects, each user shared their detections with other connected users. Once all users detected most objects in their respective areas, they turned toward the center of the room to observe all detections in the scene.

Our null hypothesis was whether the context awareness derived from object detections of a scene could be enhanced if the objects were detected using differently trained models. Guided by human input from our CoInferX component, we observed the detections and concluded that our shared perception system was capable of enhancing context awareness from multi-modal models. User 3 detected objects that Users 1 and 2 were not capable of detecting, but they were still able to observe them. If each user used a model that could detect only one type of object, Ajna would allow concurrent users to observe all detected types.

6.4 Tracking Fully Occluded Objects (Experiment 3)

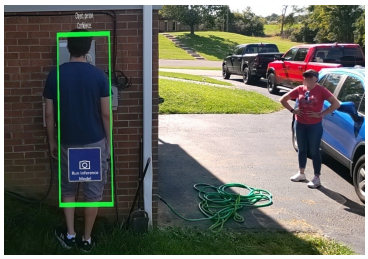
Our third experiment aimed to evaluate the tracking capability of Ajna for objects with varying degrees of occlusion, including complete occlusion. This experiment was designed to primarily test our PercepShare component. During the development and testing of Ajna, we observed that spatial anchors for a detection remained in place even when users moved away from the original detection scene. To test the tracking capability of occluded detections fully, we conducted the evaluation outdoors using three human users, each equipped with the same object detection model for persons, namely Tiny-YOLOv2.

The experiment was conducted in the vicinity of a building in the real world. User 1 was positioned 5 yards away directly facing a person, while User 2 was positioned 10 yards away around the corner of the building and observing the person who was partially occluded by the corner. User 3 was positioned on the opposite side of the building, 20 yards away, and had a completely occluded view of the person observed by User 1. Our null hypothesis was that the shared perception system would be capable of tracking partially or fully occluded detections from varying distances and angles. Figure 5 illustrates the observations made by the human users in the experiment.

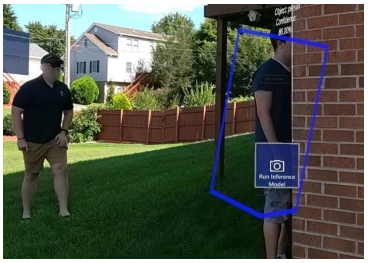
Table 2. Shared Perception Results for Occluded Detections

User	Yards from Person	Occlusion %	Observed Detection
1	5	0%	Yes
2	10	30%	Yes
2	10	80%	Yes
3	20	100%	Yes

Based on the observations and results documented in Table 2, we can accept the null hypothesis and demonstrate that detections made by Ajna can be tracked up to complete occlusion, even in outdoor settings. Our study shows that occluded detections in AR can be successful, which is particularly significant given that outdoor AR experiences are often considered subpar compared to indoor environments, as suggested by several studies [48, 54, 73]. AR experiences in outdoor environments are frequently affected by several issues, including inaccurate spatial anchors caused by the limited range of depth sensors, and poor hologram visualizations due to changes in outdoor lighting conditions, such as cloud cover. However, we found that the HoloLens 2 headset was able to generate spatial anchors outside when it was in close proximity to physical objects, such as buildings or cars. Moreover, we found that it was easier for the other connected headsets to calibrate outdoors once the host headset shared its original spatial anchors. Nevertheless, to further evaluate the design of our system and to provide a qualitative assessment, we conducted an empirical study that is detailed in the next Section.



(a) User 1 Detects Person at 86% Confidence.



(b) User 2 Receives User 1 Detection with 30% Occlusion.

(c) User 2 Tracks Detection with 80% Occlusion.

(d) User 3 Receives User 1 Detection with complete Occlusion.

Fig. 5. Occluded Object Detection Outside for Shared Perception in AR

7 USER STUDY AND RESULTS

Our IRB-approved usability study consisted of 15 participants and was designed to (1) test the feasibility of Ajna for extreme sensemaking and (2) determine its usability.

7.1 User Study Design

All participants were recruited using our department's graduate student email mailing list and using distribution lists created for our research center's mailing lists. Participants were given a written consent form for review and signature prior to any testing and were screened to ensure each participant was 18 or older. Participants were randomly grouped for each session in groups of three. Three is a common, minimum number required for a team of rescuers in a confined space [30]. Participants completed a pre-survey that recorded their gender, age, and familiarity with the research facility, AR devices, and the development of AI models/applications. Before testing, each participant was given a 10-minute tutorial on AR gestures specific to the HoloLens 2 and instructions on fitting and operating the device.

We designed the study around a mock search and rescue scenario. Search and rescue is very collaborative among a team of rescuers and aims to quickly locate a person in need of assistance [88]. As such, participants needed to work as a team to quickly locate an individual somewhere between two floors in a controlled research facility (an office building). One researcher acted as the help-seeker hidden in the building. The locations were the same for each group of participants.

The study used a within-subjects experiment design. For the first test, each participant needed to use the headset's object detection capability (using the model Tiny-YOLOv2), which created a 3-dimensional hologram bounding box around the help-seeker once identified. The test ended when all three participants found and detected the help-seeker. Then, the same three participants were asked to complete the same test without using the

AR headsets. The second test without the headsets was deliberate to gauge the difficulty of finding the help-seeker without Ajna. Each participant had to work independently as a rescuer to find the help-seeker and collaborate with each other until all three rescuers met to help the individual. During the test without AR headsets, an online audio meeting was set up so that participants could talk with each other and evaluators could record how the rescuers collaborated without the detection sharing capability through physical space.

After completion of testing, participants completed a 7-question post-survey using a 5-point Likert scale [60] focused on their perception of the performance of the prototype as it was running during their testing. The questions are presented in Section 6. In addition, we also asked participants open-ended questions to collect data on their opinions on the modular potential of different models used for Ajna. We wanted to ensure all the participants practiced and used Ajna several times so they understood how shared perception could be used.

The evaluation included 15 participants ranging in age from 19 to 44 years old, including 3 women and 12 men. Five reported never having used AR technology before, while the rest stated they had used it "once or twice" or "3 to 10" times. All were graduate students at our university with no prior search and rescue experience. Before collaborating with search and rescue professionals to test Ajna, we first wanted to sample opinions of our work in academia to gain initial feedback.

7.2 Feasibility of Ajna

We conducted a further evaluation of our prototype by assessing its ability to meet the design goals described in Section 4, as well as its potential to enhance extreme sensemaking as hypothesized. It is important to note that during the user study, two participant groups encountered unexpected technical issues with the peer-to-peer connectivity of the headsets, which resulted in the need to repeat the first test after resolving the issue and finding the help-seeker.

To investigate the impact of collaborative sensemaking, we conducted a user study to evaluate the effectiveness of Ajna in facilitating communication sharing of detections and contextual information on how to detect objects. We measured the time taken by participants to find the help-seeker, with and without using Ajna. Table 3 presents the timed results for both scenarios. The results indicate that participants took about 1 minute longer on average to find the help-seeker without using Ajna, resulting in a 15% increase in time. However, once one participant detected the help-seeker using Ajna and shared the detection with other rescuers, the remaining searchers could instantly see the location of the help-seeker, including the floor level and side.

Table 3. Timed Results (in Minutes)

	Ajna	**Control**	**Avg Decrease**
Group 1	3.43	4.09	-19.24%
Group 2	4.07	4.48	-10.07%
Group 3	8.49	9.43	-11.07%
Group 4	4.52	5.35	-18.36
Group 5	3.59	4.15	-15.6%

Table 4 and 5 present the results of a paired t-test to examine the impact of using Ajna in finding the help-seeker. Since the data was normally distributed, a significant difference (p = 0.0004) and a large effect size (d = 3.34) were observed, indicating that Ajna had a positive impact in reducing the time it took participants to find the help-seeker. It is worth noting that a delay of even one minute in responding to emergency situations can result in a 2% increase in mortality depending on the severity of the injury [71]. Thus, the potential of Ajna in improving search and rescue applications is promising.

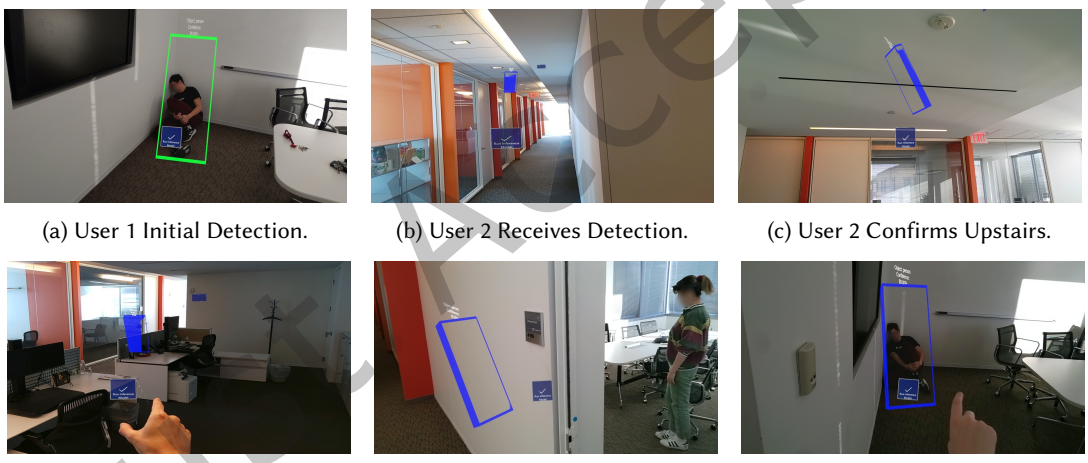
Table 4. Mean Task Completion Time (in Minutes)

	Mean	**95% CI**
Control	5.61	3.82-7.40
Ajna	4.89	3.24-6.54

Table 5. Results of Paired T-tests on Difference in Mean Task Completion Times

Metric	**t-value**	**p-value**	**LB 95% CI**	**UB 95% CI**	**Effect Size d**
Control/Ajna	-8.18	0.00044	-2.57	2.57	3.34

The ability of participants to complete our tests faster depended on their observation of the detections sent to each other, which required them to view objects through physical walls and ceilings of the office environment. Participants conveyed their opinions on this research question through our post-survey after the tests. The survey revealed that 89% of participants found this capability enhanced their situational awareness, while 83% felt more confident in locating a help-seeker while using Ajna. This indicates that humans experience an increase in situational awareness or confidence when objects of interest are tracked through solid objects. Further details on the answers from our post-survey are provided in the next subsection.



(a) User 1 Initial Detection. (b) User 2 Receives Detection. (c) User 2 Confirms Upstairs.

(d) User 3 Receives Detection. (e) User 3 Meets User 1. (f) User 2 Meets User 3.

Fig. 6. User Study Example of Mock Search & Rescue

In our user study, Figure 6 presents an example of a trial group and the various observance stages that participants followed to locate a help-seeker. Participants periodically captured images during the tests to validate the study's observations. Once one rescuer detected and sent the location of the help-seeker, the other rescuers could immediately understand where to proceed. For instance, in Figure 6b and Figure 6c, User 2 observed the bounding box of the help-seeker through the building's ceiling and instantly knew they were on a floor above. Similarly, in Figure 6e, User 3 observed the bounding box through the walls of the building and proceeded to the correct area.

7.3 The Impact of IMU Drift

The PercepShare component of Ajna relies on the embedded IMU to measure the distance each agent has moved from the mutually agreed waypoint. During our evaluation, participants noted differences in the location of the hologram sent from the participant who found the "victim" to the ones received by the other members of the group. These differences happen due to variance in the spatial measurements of the device IMUs, a phenomenon called IMU drift. This effect is discussed in more detail in other work (e.g., [92]).

Here, we seek to measure the impact this drift has on Ajna by empirically comparing the location of the original object detection to the location perceived by the other devices. For this evaluation, we used the same test site, detection locations, and distances as the original study. After sharing spatial anchors, an object detection was generated by one participant and sent to a second device at varying distances in the building. In alignment with the methodology of the original evaluation, this experiment was carried out across various floors within the building. The vertical distance was assessed within the building's architecture, which may result in a measurement slightly shorter than the building's true vertical extent. After receiving the hologram, the second participant moved to the hologram's location, as Ajna had shown them. After arriving at the location shown by the hologram, both participants measured the distance between the two holograms using an XY-coordinate grid in centimeters as a backdrop behind the detected object to determine the impact of the IMU's drift. Table 6 outlines the results, and we offer a more detailed discussion of the impact of these results in Section 8.3.

Table 6. Impact of IMU Drift

Floors Between Participants	Vertical Distance Between Participants (in meters)	Lateral Distance Between Participants (in meters)	Drift (in cm)
0	0	5	0.447
0	0	13	0.96
0	0	26	1.8
0	0	36	0.427
1	3.2	5	16.06
1	3.2	13	11.2
1	3.2	26	11.4
1	3.2	36	8.21
2	6.4	5	10.8
2	6.4	13	11.6
2	6.4	26	7.62
2	6.4	36	11.42

Based on the impact of IMU drift observed in our study, we find that the effect of drift becomes more pronounced with increasing distance between the sender and receiver of a detection. This drift can be particularly problematic in contexts where precise localization is critical. For instance, in scenarios such as firefighting in a large, multi-level building, even a small error in localization could lead to significant delays or missteps, potentially endangering lives. In these situations, the IMU drift cannot be ignored and must be addressed to ensure accurate and reliable localization.

However, a certain amount of drift might still be acceptable in contexts where vertical elevation change is the primary concern, such as distinguishing between different floors. If the drift indicates a different floor, the receiver can reasonably interpret the target detection as being on another level, which maintains a level of utility despite the drift. One potential solution is to implement a threshold-based adjustment for spatial anchors to mitigate the drift in high-precision scenarios. An additional localization adjustment could be applied when the distance between the sender and receiver exceeds a certain threshold. This adjustment could add or subtract

a calculated difference to the spatial anchor to compensate for the drift. However, this method would require further testing to determine the appropriate thresholds and calculations needed to effectively mitigate the drift based on the distance between the sender and receiver.

7.4 Usability of Ajna

To explore the usability of Ajna for assisting in sensemaking, we conducted a post-testing survey to assess the device's usability among participants. The survey consisted of the following questions, and the responses were rated on a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

(1) The search and rescue exercise was difficult without the AR headset.
(2) The search and rescue exercise was stressful while using the AR headset.
(3) The object detection capability in the AR headset helped me be aware of the situation.
(4) The system distracted me from locating the help-seeker.
(5) I felt more confident I would locate the help-seeker when wearing the AR headset.
(6) I found the system interface easy to understand and use.
(7) I liked being in control of when and when not to enable object detection.

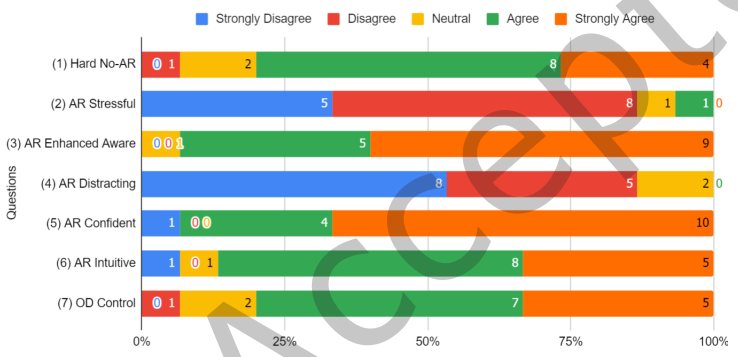


Fig. 7. Likert Questionnaire Results

The results of our post-testing survey, which aimed to investigate the usability of Ajna in supporting sensemaking, are presented in Figure 7. Notably, 93% of participants responded with "Strongly Agree" or "Agree" to Questions (3) and (5), indicating a strong belief that Ajna enhances situational awareness. Furthermore, 80% of participants responded with "Strongly Agree" or "Agree" to Question (7), suggesting a preference for a human-in-the-loop component in a shared perception system. Our findings support the conclusion that Ajna aids in object detection and enhances situational awareness. Moreover, none of the participants found the prototype to be distracting during their interactions with their partners, as indicated by the fact that no one selected "Agree" or "Strongly Agree" for Question (4). In addition to the Likert analysis, we further evaluated the participants from a System usability scale (SUS) test [10]. Table 7 summarizes the SUS scores of each participant.

The average score and its 95% confidence interval among all the participants were 81.4 ± 7.41. Based on acceptable ranges [6], 13 out of 15 participants evaluated Ajna as acceptable, while two participants (P12 and P15) evaluated it as unacceptable. Low-scored participants mainly pointed out the difficulty in interacting with the UI and learning how to use Air Tap on the device [63].

7.4.1 Observations and Feedback. After the seven survey questions, participants were also prompted with three open-ended questions. This gave participants a chance to provide additional thoughts on using Ajna for SAR applications as well as feedback on the overall study. Below are the three questions asked:

Table 7. Result of SUS Test

ID	Gender	SUS score (rank)
P01	Male	91 (A+)
P02	Male	100 (A+)
P03	Female	91 (A+)
P04	Male	91 (A+)
P05	Male	86 (A+)
P06	Female	86 (A+)
P07	Female	74 (B)
P08	Male	80 (A-)
P09	Male	74 (B)
P10	Male	80 (A-)
P11	Male	88.5 (A+)
P12	Male	68.5 (C)
P13	Male	97 (A+)
P14	Male	74 (B)
P15	Male	40 (F)

(1) What did you like about the experiences?
(2) What could be improved about the Shared Perception software you used during this study?
(3) Are there any other applications of this technology (besides search and rescue) that you think could be helpful?

Users preferred using Ajna over nothing to assist in sensemaking. Thirteen out of the 15 participants provided positive feedback regarding their preference for using Ajna over not for the first question. (The other two participants did not answer the question.) Some felt it was the best way to test the system, while only a couple of participants thought the system needed to be improved more. P04 perceived the ability of Ajna to smoothly track occluded people as very novel, stating, *"The ability to see locations of detections sent from my team through the ceiling was amazing!"* Some participants noted that this capability greatly assisted in sensemaking in unknown environments. For example, P05 stated, *"We thought it was very useful for finding people when they don't know the building environment,"* and P14 said, *"It is a very efficient solution for search and rescue in an unknown territory."* Other participants noted that using Ajna with others was critical in more efficient collaboration compared to without the headset, such as P07, who stated, *I liked the enhanced intelligence that using the AR device provided me in the search and rescue operation. Having a device to fall back on for critical information that can be shared with peers proved helpful in specific scenarios where I myself got lost in the mission."*

Users wanted enhanced communication and a smoother UI. For the second question, 14 out of the 15 participants provided constructive feedback on how to improve Ajna for a more intuitive experience. Most feedback was related to further enhancement in how communication is relayed in the peer-to-peer network. Participants thought the exercise could have been smoother if Ajna assisted in sensemaking for contexts where reliable internet could be available, such as urban environments. For example, P04 stated, *"The connections seemed to be unstable sometimes between the headsets. Maybe use the cloud instead of a peer-to-peer network?"* P06 stated, *"I think stability should be improved,"* and P14 said, *"User experience can be improved by making software smooth."* Other improvement ideas were related to revising the UI and replacing the Air Tap interactivity with button pressing instead. P07, P11, and P13 all provided the same idea of some kind of notification to pop up when a fellow rescuer sent a detection for others to see. P07 said:

> *It would certainly help to inform all peers when a new target of interest has been identified and marked by a peer. Peers don't get to know that such a new object has been identified. This has the potential to speed up intelligence sharing and could help optimize SAR operations by other peers involved in the current mission.*

While P11 suggested that the notification be some kind of audio sound instead of a visual notification, *"An audio ping should be played when a fellow rescuer sends us a detection."* And P13 stated *"There could be a notification marker that automatically tells all the users that the target has been detected and marked."* Some users had difficulty using Air Tap as a means to interact with the UI, such as P10, who stated, *"Mainly the UI should be improved. Especially the UI to broadcast the location to others using Air Tap."* And P15, who stated, *"The pinch functionality to mark the target was difficult to use."*

Users proposed other applications on how Ajna could be used to help others. Ten of the 15 participants provided additional application ideas for which Ajna could be used. Five participants chose not to answer. Most ideas given were different variations of Ajna that could be used for interactive gaming; for example, P05 mentioned Pokemon Go as an inspiration, and P13 suggested tourism applications. Other ideas related to use for industries or activities similar to our original ideas. For example, P04 mentioned law enforcement, and P14 said, *"It can be used to identify objects or avoid barriers for people with a disability. Also, it can help military defense by predicting an enemy soldier's movement."* P06 and P11 thought of other practical use cases for manufacturing and civil infrastructure repairs. P11 elaborated, "If the location and data of each component are saved on Ajna, it can be used for detecting objects behind the wall that require repair, for example, pipes." P06 said:

> *Ajna would be perfect for smart warehouses or Internet of Things use cases. The synchronized devices of multiple users could help increase the productivity of warehouse workers or factory technicians to keep a tab on certain time-critical operations that need constant supervision and corrective action.*

8 DISCUSSION

In the first evaluation study, we conducted three system experiments on Ajna's capability in extreme sensemaking with object detection. Based on the quantitative results and user observations, we summarize the findings by usability, future applications, and notable limitations.

8.1 Usefulness of Ajna

In the second evaluation study, we confirmed that Ajna enabled participants physically spread throughout a large multi-floor building to efficiently track objects of interest faster than without the system. Quantitative and qualitative feedback suggests that the users also gained confidence and trust in Ajna to quickly navigate to a victim of a mock SAR once a detection was distributed from a fellow rescuer. The results from our research validate similar results found in a previous study by Fisher et al. [29]. Their work highlighted that distributed sensemaking, enabled by shared mental models, communication networks, and the use of digital technologies, improved a team's ability to interpret complex situations, coordinate actions, and adapt to dynamic environments. The authors evaluated their work through a qualitative research approach similar to ours but investigated the interactions and sensemaking processes. Ultimately, they concluded that distributed cognition provides a mechanism for leveraging collective intelligence and expertise, leading to more effective sensemaking and decision-making in high-pressure contexts.

While their research focused on interactions and sensemaking processes, without the inclusion of AR or AI platforms in their collaborative technology, we can infer similar usability enhancements from Ajna based on our Likert results and initial System Usability Scale (SUS) scores. Our evaluation indicates that Ajna's overall usability metric surpasses nominal levels, with the PercepShare component receiving particularly positive feedback for its

role in sensemaking and knowledge distribution among rescuers, even in scenarios with complete occlusion. We rank our sensemaking application on the HoloLens 2 against others in similar problem spaces in Table 8.

Table 8. Sensemaking Applications for AR

Application	Year	Device	IMU	Scalable	HITL	AI-Integrated	Self-Contained
BARS [43]	2000	Glasstron	✓				
Rapid Assessment [45]	2007	Emulator	✓				
The Prototype [38]	2019	HoloLens	✓			✓	
Inspector Assistant [46]	2019	HoloLens	✓			✓	✓
MARLIN [4]	2019	Smartphone				✓	✓
CollabAR [53]	2021	Magic Leap				✓	
SAVE [58]	2022	HoloLens	✓				✓
WIRMS [85]	2022	HoloLens	✓			✓	✓
DeepMix [34]	2022	HoloLens	✓			✓	
X-AR [9]	2023	HoloLens	✓				
IVAS [84]	2023	HoloLens	✓	✓	✓		
MARS [47]	2023	HoloLens	✓	✓	✓		
Live Tracker [12]	2023	HoloLens	✓	✓	✓		
Ajna	**2024**	**HoloLens**	**✓**	**✓**	**✓**	**✓**	**✓**

We conclude that Ajna advances usability in intelligent systems for extreme sensemaking scenarios. All applications listed in Table 8 are wearable sensemaking applications for AR except the "Rapid Assessment" and "MARLIN" tools. Defined as scalable, Ajna supports increasing numbers of concurrent users, thereby amplifying the collective cognitive process. As an AI-centric platform, it utilizes Inertial Measurement Unit (IMU) data in combination with artificial intelligence to deepen distributed cognition. This system is designed to be self-contained and function independently without requiring additional equipment or external support. Additionally, Ajna incorporates a human-in-the-loop (HITL) approach, allowing users to interactively validate or communicate AI-driven findings. This integration not only enhances the accuracy and effectiveness of the sensemaking process but also ensures that the system remains adaptable and responsive to real-world user interactions and environments.

8.2 Broader Applications

Our shared perception system offers numerous potential applications beyond the search and rescue scenarios explored in this study. One potential application is in the field of physical security or law enforcement, where the system could aid in stealthy infiltration of indoor spaces. For instance, some officers could use Ajna to detect and locate the positions of hostile targets within a building, create bounding boxes around them, and then share that information with other officers about to enter the building from the other side, thus increasing their situational awareness and reducing the risk of injury or ambush.

Another potential application is in wildlife detection and tracking, where Ajna could provide a unique platform for enhancing situational awareness for safari explorers or park rangers. Multiple explorers or rangers could spread out in a wildlife area, each using a headset to detect different species of animals and then share those detections with others through the shared perception system. This way, all users of the system could quickly locate animals in the area, potentially reducing the time and effort needed to track and monitor wildlife.

Moreover, our shared perception system could have applications in disaster response, where multiple search and rescue teams could use the system to coordinate their efforts and share information about the location and status of survivors. It could also be useful in infrastructure inspection or maintenance, where multiple inspectors could use the system to detect and locate damaged or malfunctioning equipment, share that information with others, and facilitate more efficient and accurate repairs. Water treatment systems may be a good example of how to test Ajna for civil infrastructure applications. Several AI models exist to detect corrosion or cracks in pipes [49, 72]. However, AR has only been used so far to enhance the information available to repair workers such as overlaying real-time sensor data, like water flow rates, pressure levels, or temperature readings, onto the physical infrastructure [68]. No examples have been found that use AI as well for these AR contexts.

Overall, the potential applications of our wearable shared perception system are numerous, and the system could be adapted to various contexts and fields, making it a promising platform for enhancing sensemaking in a variety of settings with extreme conditions. Further research and testing are needed to explore the feasibility and effectiveness of using Ajna in these contexts.

8.3 Limitations and Future Research

Our research with Ajna revealed several limitations impacting the user experience and effectiveness of the system. These limitations, along with potential solutions and future research directions, are detailed below. Our future efforts will focus on adopting and developing robust AI assurance methods. This approach aims to enhance the reliability, transparency, and explainability of Ajna's AI-driven functionalities, fostering greater confidence and trust among users. Advancing these methods will also contribute to more predictable and understandable AI behaviors, addressing concerns highlighted in our findings.

- **Object Detection Model Constraints:** The object detection models used in Ajna, particularly on the HoloLens 2, showed limitations related to model size. Larger models were found to negatively impact the frames per second of detection and increase system latency, affecting the overall user experience. We recommend using smaller models, such as Tiny-YOLO, for optimal performance when edge or cloud computations are not feasible. Future research could explore more efficient object detection models for AR environments.
- **Depth Estimation Accuracy:** Depth estimation for creating three-dimensional bounding boxes sometimes suffered from accuracy issues, leading to reduced confidence among users. This was partly due to how the depth sensor array interacts with objects and the user's perspective. As a workaround, we hard-coded default depths for certain object classifications, but more systematic improvements are needed. Future research should focus on enhancing the accuracy of depth estimation in AR.
- **Elevation Change Detection with HoloLens 2:** As shown in Section 7.3, the measured spatial drift between object detection generally increased as the distance and elevation changed between evaluation participants. At its maximum, this results in only between 11 cm and 16 cm of drift between detections. These differences may be attributed to the physical gyroscope in the HoloLens 2, which requires time to adjust to changes in vertical elevation and the natural drift noted in IMU measurement over distance [19, 44, 92]. Given the outsized impact of the elevation changes specifically, one solution could be to add an altimeter to the prototype to aid in the estimation of vertical elevation change. However, as a prototype designed to explore the feasibility of extreme sensemaking in AR, we leave this to future work.
- **Challenges in Outdoor Settings:** Experiments in outdoor settings highlighted difficulties with spatial mapping and object detection due to variable lighting conditions and the need for physical reference objects [48, 54, 73]. Improving the performance of the HoloLens sensors in such environments is a key area for future research.

To summarize and validate our findings, our future research will focus on two areas to enhance our shared perception prototype:

(1) **Comprehensive Collaborative Sensemaking Study:** We plan a larger user study to evaluate collaborative sensemaking with a significant Human-in-the-Loop (HITL) component. This involves extending our network of connected HoloLens 2 headsets for more concurrent users, allowing us to test crowdsourced methodologies for managing shared detections and analyzing information flow and collaboration patterns through network analysis. Incorporating context-driven AI could improve the relevancy of information in various environments, especially in minimizing less pertinent outdoor settings. By tailoring AI responses to specific contexts, it may ensure that the system's insights are directly applicable to the users' immediate circumstances. Collaboration with search and rescue professionals will also offer practical insights into Ajna's application, ensuring the system's design and functionality meet the nuanced demands of real-world rescue scenarios.
(2) **System Improvement with AI Assurance:** The opaque "black box" decision-making of AI models like YOLO challenges user trust and understanding, highlighted by frustrations despite human-AI collaboration enhancements [41]. Addressing this, our future direction includes developing XAI components for AR object detection to foster user interaction and trust in Ajna. Concurrently, we plan to refine the system by adjusting detection confidence levels through user feedback while managing multiple inputs for identical detections. This strategy will likely explore existing and new AI Assurance principles to improve the system's trustworthiness and explainability for shared perception systems [7].

While Ajna has demonstrated its potential in enhancing extreme sensemaking, these limitations and future research directions highlight the need for ongoing development to refine and expand the system's capabilities. Addressing these challenges will be crucial for advancing Ajna's effectiveness and broadening its applicability across various scenarios.

9 CONCLUSION

Extreme sensemaking is the act of sensemaking performed under extreme conditions, such as navigating a partially collapsed building post-earthquake, requiring quick decision-making for locating survivors. This paper presents Ajna, a novel shared perception system for enhancing extreme sensemaking through collaborative object detection by leveraging the expertise of a human-in-the-loop. Our paper offers the following key findings and contributions. First, Ajna leverages shared spatial AR references with minimal drift without relying on QR codes, additional edge hardware, or cloud services, ensuring smooth tracking of occluded objects. Second, a user study-based evaluation of Ajna with 15 participants in a simulated search and rescue environment demonstrated a significant decrease in task completion times; i.e., a 15% reduction in victim locating time. Third, the user study also showed that 89% of participants reported enhanced situational awareness when using Ajna, and 13 out of 15 participants found Ajna acceptable in terms of usability.

Ajna represents an AR technology advancement for extreme sensemaking, offering promising applications in search and rescue operations, law enforcement training, and complex navigation tasks. These findings expand the boundaries of interactive intelligent systems and introduce exciting possibilities for applying our novel approach across diverse domains.

ACKNOWLEDGMENTS

This research would not have been possible without funding assistance from the U.S. Army Combat Capabilities Development Command. We also thank Matthew Corbett for assisting in the development of the Ajna prototype, along with Dr. Bo Ji for his guidance in this work. Finally, a word of thanks to Yingjie Wang and Ryan Hankard

for their input and feedback in the development of this paper, as well as the team at the A3 Lab at Virginia Tech (https://ai.bse.vt.edu/).

REFERENCES

[1] Luc Claesen Puxun Tu Jan Egger Abel J Lungu, Wout Swinkels and Xiaojun Chen. 2021. A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: an extension to different kinds of surgery. *Expert Review of Medical Devices* 18, 1 (2021), 47–62. https://doi.org/10.1080/17434440.2021.1860750 arXiv:https://doi.org/10.1080/17434440.2021.1860750 PMID: 33283563.

[2] Sophia J. Abraham, Zachariah Carmichael, Sreya Banerjee, Rosaura G. VidalMata, Ankit Agrawal, Md Nafee Al Islam, Walter J. Scheirer, and Jane Cleland-Huang. 2021. *Adaptive Autonomy in Human-on-the-Loop Vision-Based Robotics Systems.* Vol. abs/2103.15053. https://conf.researchr.org/details/wain-2021/wain-2021-papers/11/Adaptive-Autonomy-in-Human-on-the-Loop-Vision-Based-Robotics-Systems

[3] Adel, Liangkai Zhang, Jianing Wei, Artsiom Ablavatski, and Matthias Grundmann. 2020. Objectron: A Large Scale Dataset of Object-Centric Videos in the Wild with Pose Annotations. *CoRR* abs/2012.09988 (2020). arXiv:2012.09988 https://arxiv.org/abs/2012.09988

[4] Kittipat Apicharttrisorn, Xukan Ran, Jiasi Chen, Srikanth V. Krishnamurthy, and Amit K. Roy-Chowdhury. 2019. Frugal following: power thrifty object detection and tracking for mobile augmented reality. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems* (New York, New York) *(SenSys '19)*. Association for Computing Machinery, New York, NY, USA, 96–109. https://doi.org/10.1145/3356250.3360044

[5] Ronald T. Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6, 4 (Aug. 1997), 355–385. https://doi.org/10.1162/pres.1997.6.4.355 _eprint: https://direct.mit.edu/pvar/article-pdf/6/4/355/1623026/pres.1997.6.4.355.pdf.

[6] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *J. Usab. Stud.* 4, 3 (2009), 114–123.

[7] Feras A. Batarseh, Laura Freeman, and Chih-Hao Huang. 2021. A survey on artificial intelligence assurance. *Journal of Big Data* 8, 1 (26 Apr 2021), 60. https://doi.org/10.1186/s40537-021-00445-7

[8] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. . https://doi.org/10.1561/1100000049

[9] Tara Boroushaki, Maisy Lam, Laura Dodds, Aline Eid, and Fadel Adib. 2023. Augmenting Augmented Reality with Non-Line-of-Sight Perception. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. USENIX Association, Boston, MA, 1341–1358.

[10] John Brooke. 1996. SUS: A "quick and dirty" usability. *Usab. Eval. Industr.* (1996), 189–194.

[11] Julie Carmigniani, Borko Furht, Marco Anisetti, Paolo Ceravolo, Ernesto Damiani, and Misa Ivkovic. 2011. Augmented reality technologies, systems and applications. *Multimedia Tools and Applications* 51, 1 (Jan 2011), 341–377. https://doi.org/10.1007/s11042-010-0660-6

[12] Theodoros Chalimas and Katerina Mania. 2023. Cross-Device Augmented Reality for Fire and Rescue Operations based on Thermal Imaging and Live Tracking. In *Proceedings of the 1st Joint Workshop on Cross Reality (JWCR23) at ISMAR 2023*. Sydney, Australia. https://cross-realities.org/proceedings/JWCR23_paper_11.pdf

[13] Jongin Choe and Sanghyun Seo. 2020. A 3D Real Object Recognition and Localization on SLAM based Augmented Reality Environment. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*. 745–746. https://doi.org/10.1109/CSCI51800.2020.00140

[14] Christopher B Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 2016. 3D-R2N2: A unified approach for single and multi-view 3D object reconstruction. In *European Conference on Computer Vision*. Springer, Cham, 628–644.

[15] Matthew Corbett, Brendan David-John, Jiacheng Shang, Y. Charlie Hu, and Bo Ji. 2023. BystandAR: Protecting Bystander Visual Data in Augmented Reality Systems. In *21st Annual International Conference on Mobile Systems, Applications and Services (MobiSys '23)*. https://doi.org/10.1145/3581791.3596830

[16] Alessandra Corneli, Berardo Naticchia, Alessandro Cabonari, and Frédéric Bosché. 2019. Augmented Reality and Deep Learning towards the Management of Secondary Building Assets. In *Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC)*, Mohamed Al-Hussein (Ed.). International Association for Automation and Robotics in Construction (IAARC), Banff, Canada, 332–339. https://doi.org/10.22260/ISARC2019/0045

[17] Microsoft Corportation. 2021. *Azure Spatial Anchors overview.* Retrieved January 10, 2022 from https://docs.microsoft.com/en-us/azure/spatial-anchors/overview

[18] António Correia, Andrea Grover, Daniel Schneider, Ana Paula Pimentel, Ramon Chaves, Marcos Antonio de Almeida, and Benjamim Fonseca. 2023. Designing for Hybrid Intelligence: A Taxonomy and Survey of Crowd-Machine Interaction. *Applied Sciences* 13, 4 (2023). https://doi.org/10.3390/app13042198

[19] Gabriel M. Costa, Marcelo R. Petry, João G. Martins, and António Paulo G. M. Moreira. 2024. Assessment of Multiple Fiducial Marker Trackers on Hololens 2. *IEEE Access* 12 (2024), 14211–14226. https://doi.org/10.1109/ACCESS.2024.3356722

[20] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5828–5839.

[21] Jared Van Dam, Alexander Krasner, and Joseph L. Gabbard. 2020. Augmented Reality for Infrastructure Inspection with Semi-autonomous Aerial Systems: An Examination of User Performance, Workload, and System Trust. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 742–743. https://doi.org/10.1109/VRW50115.2020.00222

[22] Archi Dasgupta, Mark Manuel, Rifat Sabbir Mansur, Nabil Nowak, and Denis Gračanin. 2020. Towards Real Time Object Recognition For Context Awareness in Mixed Reality: A Machine Learning Approach. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 262–268. https://doi.org/10.1109/VRW50115.2020.00054

[23] Hitesh Dhiman, Sascha Martinez, Volker Paelke, and Carsten Röcker. 2018. Head-Mounted Displays in Industrial AR-Applications: Ready for Prime Time?. In *HCI in Business, Government, and Organizations*, Fiona Fui-Hoon Nah and Bo Sophia Xiao (Eds.). Springer International Publishing, Cham, 67–78.

[24] Ruofei Du, Eric Turner, Maksym Dzitsiuk, Luca Prasso, Ivo Duarte, Jason Dourgarian, Joao Afonso, Jose Pascoal, Josh Gladstone, Nuno Cruces, Shahram Izadi, Adarsh Kowdle, Konstantine Tsotsos, and David Kim. 2020. DepthLab: Real-time 3D Interaction with Depth Maps for Mobile Augmented Reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 829–843. https://doi.org/10.1145/3379337.3415881

[25] Thomas Joseph Duffy. 2016. *Collaborative sensemaking*. Ph. D. Dissertation. University of Birmingham.

[26] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2019. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies* 131 (2019), 81–98. https://doi.org/10.1016/j.ijhcs.2019.05.011 50 years of the International Journal of Human-Computer Studies. Reflections on the past, present and future of human-centred technologies.

[27] Giuliano Ferreira Dela Coleta, Alexandre Cardoso, Edgard Afonso Lamounier Júnior, and Gerson Flávio Mendes de Lima. 2019. Telecommunications Field Operations Supported by Augmented Reality – a Systematic Review. In *2019 21st Symposium on Virtual and Augmented Reality (SVR)*. 77–83. https://doi.org/10.1109/SVR.2019.00028

[28] Stephen M. Fiore, Travis J. Wiltshire, Rachel A. Lashlee, and Eduardo Salas. 2010. Distributed sensemaking: A case study of military analysis. *Journal of Organizational Behavior* 31, 2-3 (2010), 291–307. https://doi.org/10.1002/job.619

[29] Kristie Fisher, Scott Counts, and Aniket Kittur. 2012. Distributed sensemaking: improving sensemaking by leveraging the efforts of previous users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 247–256. https://doi.org/10.1145/2207676.2207711

[30] Canadian Centre for Occupational Health and (CCOHS) Safety. 2022. *Confined Space - Program*. Canadian Centre for Occupational Health and Safety, (CCOHS). Retrieved Dec 7, 2022 from https://www.ccohs.ca/oshanswers/hsprograms/confinedspace_program.html

[31] Marlena R. Fraune, Ahmed S. Khalaf, Mahlet Zemedie, Poom Pianpak, Zahra NaminiMianji, Sultan A. Alharthi, Igor Dolgov, Bill Hamilton, Son Tran, and Z.O. Toups. 2021. Developing Future Wearable Interfaces for Human-Drone Teams through a Virtual Drone Search Game. *International Journal of Human-Computer Studies* 147 (2021), 102573. https://doi.org/10.1016/j.ijhcs.2020.102573

[32] Andreas Geiger, Philip Lenz, and Raquel Urtasun. 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 3354–3361. https://doi.org/10.1109/CVPR.2012.6248074

[33] Yalda Ghasemi, Heejin Jeong, Sung Ho Choi, Kyeong-Beom Park, and Jae Yeol Lee. 2022. Deep learning-based object detection in augmented reality: A systematic review. *Computers in Industry* 139 (2022), 103661. https://doi.org/10.1016/j.compind.2022.103661

[34] Yongjie Guan, Xueyu Hou, Nan Wu, Bo Han, and Tao Han. 2022. DeepMix: Mobility-Aware, Lightweight, and Hybrid 3D Object Detection for Headsets. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services* (Portland, Oregon) *(MobiSys '22)*. Association for Computing Machinery, New York, NY, USA, 28–41. https://doi.org/10.1145/3498361.3538945

[35] Klaus Haller. 2022. *Quality Assurance in and for AI*. Apress, Berkeley, CA, 61–83. https://doi.org/10.1007/978-1-4842-7824-6_3

[36] Xu Han, Ying Chen, Qinna Feng, and Heng Luo. 2022. Augmented Reality in Professional Training: A Review of the Literature from 2001 to 2020. *Applied Sciences* 12, 3 (2022). https://doi.org/10.3390/app12031024

[37] Andreas Holzinger, Anna Saranti, Anne-Christin Hauschild, Jacqueline Beinecke, Dominik Heider, Richard Roettger, Heimo Mueller, Jan Baumbach, and Bastian Pfeifer. 2023. Human-in-the-Loop Integration with Domain-Knowledge Graphs for Explainable Federated Deep Learning. In *Machine Learning and Knowledge Extraction*, Andreas Holzinger, Peter Kieseberg, Federico Cabitza, Andrea Campagner, A. Min Tjoa, and Edgar Weippl (Eds.). Springer Nature Switzerland, Cham, 45–64.

[38] Burkhard Hoppenstedt, Klaus Kammerer, Manfred Reichert, Myra Spiliopoulou, and Rüdiger Pryss. 2019. *Convolutional Neural Networks for Image Recognition in Mixed Reality Using Voice Command Labeling*. Number 11614 in Lecture Notes in Computer Science. Springer, 63–70. http://dbis.eprints.uni-ulm.de/1764/

[39] Mingwei Hu, Dongdong Weng, Feng Chen, and Yongtian Wang. 2020. Object Detecting Augmented Reality System. In *2020 IEEE 20th International Conference on Communication Technology (ICCT)*. 1432–1438. https://doi.org/10.1109/ICCT50939.2020.9295761

[40] Yucheng Hu, Zhonghong Ou, Xiangyu Xu, and Meina Song. 2019. *A Crowdsourcing Repeated Annotations System for Visual Object Detection*. Association for Computing Machinery, New York, NY, USA. https://doi-org.ezproxy.lib.vt.edu/10.1145/3387168.3387242

[41] Fatima Hussain, Rasheed Hussain, and Ekram Hossain. 2021. Explainable artificial intelligence (XAI): An engineering perspective. *arXiv preprint arXiv:2101.03613* (2021).

[42] Janna Huuskonen and Timo Oksanen. 2019. Augmented Reality for Supervising Multirobot System in Agricultural Field Operation. *IFAC-PapersOnLine* 52, 30 (2019), 367–372. https://doi.org/10.1016/j.ifacol.2019.12.568 6th IFAC Conference on Sensing, Control and Automation Technologies for Agriculture AGRICONTROL 2019.

[43] Simon Julier, Yohan Baillot, Marco Lanzagorta, Dennis Brown, and Lawrence Rosenblum. 2001. *BARS: Battlefield Augmented Reality System.* Technical Report ADP010892. NAVAL RESEARCH LAB WASHINGTON DC ADVANCED INFORMATION TECHNOLOGY. 8.0 pages. https://apps.dtic.mil/sti/citations/ADP010892 APPROVED FOR PUBLIC RELEASE.

[44] David Jurado-Rodríguez, Rafael Muñoz-Salinas, Sergio Garrido-Jurado, and Rafael Medina-Carnicer. 2021. Design, Detection, and Tracking of Customized Fiducial Markers. *IEEE Access* 9 (2021), 140066–140078. https://doi.org/10.1109/ACCESS.2021.3118049

[45] Vineet R. Kamat and Sherif El-Tawil. 2007. Evaluation of Augmented Reality for Rapid Assessment of Earthquake-Induced Building Damage. *Journal of Computing in Civil Engineering* 21, 5 (2007), 303–310. https://doi.org/10.1061/(ASCE)0887-3801(2007)21:5(303)

[46] Enes Karaaslan, Ulas Bagci, and Fikret Necati Catbas. 2019. Artificial Intelligence Assisted Infrastructure Assessment using Mixed Reality Systems. *Transportation Research Record* 2673, 12 (2019), 413–424. https://doi.org/10.1177/0361198119839988 arXiv:https://doi.org/10.1177/0361198119839988

[47] J. Keller. 2023. *China's military unveils heads-up display to let soldiers shoot around corners: Meet the MARS, the Chinese military's IVAS clone.* Task & Purpose. https://taskandpurpose.com/tech-tactics/china-military-augmented-reality-system-weapons/

[48] Steven J. Kerr, Mark D. Rice, Yinquan Teo, Marcus Wan, Yian Ling Cheong, Jamie Ng, Lillian Ng-Thamrin, Thant Thura-Myo, and Dominic Wren. 2011. Wearable Mobile Augmented Reality: Evaluating Outdoor User Experience *(VRCAI '11)*. Association for Computing Machinery, New York, NY, USA, 209–216. https://doi.org/10.1145/2087756.2087786

[49] Saffeer M. Khan, Syed A. Haider, and Ishaq Unwala. 2020. A Deep Learning Based Classifier for Crack Detection with Robots in Underground Pipes. In *2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET)*. 78–81. https://doi.org/10.1109/HONET50430.2020.9322665

[50] Jean Lahoud and Bernard Ghanem. 2017. 2D-Driven 3D Object Detection in RGB-D Images. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 4632–4640. https://doi.org/10.1109/ICCV.2017.495

[51] Ze-Hao Lai, Wenjin Tao, Ming C. Leu, and Zhaozheng Yin. 2020. Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing. *Journal of Manufacturing Systems* 55 (2020), 69–81. https://doi.org/10.1016/j.jmsy.2020.02.010

[52] Jean-François Lalonde. 2018. Deep Learning for Augmented Reality. In *2018 17th Workshop on Information Optics (WIO)*. 1–3. https://doi.org/10.1109/WIO.2018.8643463

[53] Guohao Lan, Zida Liu, Yunfan Zhang, Tim Scargill, Jovan Stojkovic, Carlee Joe-Wong, and Maria Gorlatova. 2021. Edge-Assisted Collaborative Image Recognition for Mobile Augmented Reality. *ACM Trans. Sen. Netw.* 18, 1, Article 9 (oct 2021), 31 pages. https://doi.org/10.1145/3469033

[54] Yuan Li, Ibrahim A. Tahmid, Feiyu Lu, and Doug A. Bowman. 2022. Evaluation of Pointing Ray Techniques for Distant Object Referencing in Model-Free Outdoor Collaborative Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* (2022), 1–11. https://doi.org/10.1109/TVCG.2022.3203094

[55] Luyang Liu, Hongyu Li, and Marco Gruteser. 2019. Edge Assisted Real-Time Object Detection for Mobile Augmented Reality. In *The 25th Annual International Conference on Mobile Computing and Networking* (Los Cabos, Mexico) *(MobiCom '19)*. Association for Computing Machinery, New York, NY, USA, Article 25, 16 pages. https://doi.org/10.1145/3300061.3300116

[56] Zimo Liu, Jingya Wang, Shaogang Gong, Dacheng Tao, and Huchuan Lu. 2019. *Deep Reinforcement Active Learning for Human-in-the-Loop Person Re-Identification.* IEEE, 6121–6130. https://doi.org/10.1109/ICCV.2019.00622

[57] Mark A. Livingston, Lawrence J. Rosenblum, Simon J. Julier, Dennis Brown, Yohan Baillot, J. Edward Swan II, Joseph L. Gabbard, and Deborah Hix. 2002. An Augmented Reality System for Military Operations in Urban Terrain. In *Proceedings of the Interservice / Industry Training, Simulation, & Education Conference (I/ITSEC '02)*. Orlando, FL.

[58] John Luksas, Kelsey Quinn, Joseph L. Gabbard, Mariam Hasan, Janet He, Neha Surana, Moustafa Tabbarah, and Nishant Kishan Teckchandani. 2022. Search and Rescue AR Visualization Environment (SAVE): Designing an AR Application for Use with Search and Rescue Personnel. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 488–492. https://doi.org/10.1109/VRW55335.2022.00109

[59] Cristina Manresa-Yee, Silvia Ramis, F. Xavier Gaya-Morey, and Jose M. Buades. 2024. Impact of Explanations for Trustworthy and Transparent Artificial Intelligence. In *Proceedings of the XXIII International Conference on Human-Computer Interaction* (Lleida, Spain) *(Interaccion '23)*. Association for Computing Machinery, New York, NY, USA, Article 14, 8 pages. https://doi.org/10.1145/3612783.3612798

[60] S. A. McLeod. 2019. Likert Scale Definition, Examples and Analysis. *Simply Psychology* (2019). https://www.simplypsychology.org/likert-scale.html

[61] Microsoft. [n. d.]. Mixed Reality Design. https://learn.microsoft.com/en-us/windows/mixed-reality/design/design?culture=en-us&country=us. Accessed on June 4, 2023.
[62] Microsoft. 2022. *CameraIntrinsics Class Clas.* Microsoft. Retrieved Nov 20, 2022 from https://learn.microsoft.com/en-us/uwp/api/windows.media.devices.core.cameraintrinsics?view=winrt-22621
[63] Microsoft. 2022. *HoloLens 2 gestures for navigating a guide in Dynamics 365 Guides.* Microsoft. Retrieved Nov 20, 2022 from https://learn.microsoft.com/en-us/dynamics365/mixed-reality/guides/operator-gestures-hl2
[64] Microsoft. 2022. *Spatial awareness getting started — MRTK2.* Microsoft. Retrieved Nov 6, 2022 from https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/features/spatial-awareness/spatial-awareness-getting-started?view=mrtkunity-2022-05
[65] Microsoft. 2022. *What is Mixed Reality Toolkit 2?* Microsoft. Retrieved Oct 13, 2022 from https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/?view=mrtkunity-2022-05
[66] Microsoft. 2022. *What's a Universal Windows Platform (UWP) app?* Microsoft. Retrieved Nov 20, 2022 from https://learn.microsoft.com/en-us/windows/uwp/get-started/universal-application-platform-guide
[67] Microsoft. 2022. *XRAnchorTransferBatch Clas.* Microsoft. Retrieved Nov 20, 2022 from https://learn.microsoft.com/en-us/dotnet/api/microsoft.mixedreality.openxr.xranchortransferbatch?view=mixedreality-openxr-plugin-1.6
[68] Domenica Mirauda, Ugo Erra, Roberto Agatiello, and Marco Cerverizzo. 2017. Applications of Mobile Augmented Reality to Water Resources Management. *Water* 9, 9 (2017). https://www.mdpi.com/2073-4441/9/9/699
[69] Niluthpol Chowdhury Mithun, Kshitij S. Minhas, Han-Pang Chiu, Taragay Oskiper, Mikhail Sizintsev, Supun Samarasekera, and Rakesh Kumar. 2023. Cross-View Visual Geo-Localization for Outdoor Augmented Reality. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR).* 493–502. https://doi.org/10.1109/VR55154.2023.00064
[70] Eva Mohedano, Kevin McGuinness, Graham Healy, Noel E. O'Connor, Alan F. Smeaton, Amaia Salvador, Sergi Porta, and Xavier Giró-i Nieto. 2015. Exploring EEG for Object Detection and Retrieval. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2671188.2749368
[71] Ahmed A.H. Nasser, Charlie Nederpelt, Majed El Hechi, April Mendoza, Noelle Saillant, Peter Fagenholz, George Velmahos, and Haytham M.A. Kaafarani. 2020. Every minute counts: The impact of pre-hospital response time and scene time on mortality of penetrating trauma patients. *The American Journal of Surgery* 220, 1 (2020), 240–244. https://doi.org/10.1016/j.amjsurg.2019.11.018
[72] Peter Oyekola, Kolawole Somade, Shoeb Syed, and Owen Apis. 2021. Application of Computer Vision in Pipeline Inspection Robot. https://doi.org/10.46254/AN11.20210374
[73] Chris Panou, Lemonia Ragia, Despoina Dimelli, and Katerina Mania. 2018. An Architecture for Mobile Outdoors Augmented Reality for Cultural Heritage. *ISPRS International Journal of Geo-Information* 7, 12 (2018). https://doi.org/10.3390/ijgi7120463
[74] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. 2017. PointNet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 652–660.
[75] Qwake Technologies. [n. d.]. *Qwake-Technologies C-thru.* https://www.qwake.tech/
[76] Rudraksha Ratna. [n. d.]. Ajna Chakra. https://www.rudraksha-ratna.com/articles/ajna-chakra. Accessed on June 4, 2023.
[77] Orod Razeghi. 2015. *An investigation of a human in the loop approach to object recognition.* Ph. D. Dissertation. University of Nottingham.
[78] Jeba Rezwana and Mary Lou Maher. 2023. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 67 (sep 2023), 28 pages. https://doi.org/10.1145/3519026
[79] M. L. Yuan S. K. Ong and A. Y. C. Nee. 2008. Augmented reality applications in manufacturing: a survey. *International Journal of Production Research* 46, 10 (2008), 2707–2742. https://doi.org/10.1080/00207540601064773 arXiv:https://doi.org/10.1080/00207540601064773
[80] Kate Saenko and Trevor Darrell. 2007. *Object Category Recognition Using Probabilistic Fusion of Speech and Image Classifiers.* Springer-Verlag, Berlin, Heidelberg. https://dl.acm.org/doi/abs/10.5555/1787422.1787428
[81] Aron Schmied, Tobias Fischer, Martin Danelljan, Marc Pollefeys, and Fisher Yu. 2023. R3D3: Dense 3D Reconstruction of Dynamic Scenes from Multiple Cameras. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).* 3216–3226.
[82] Julian Schuir, René Brinkhege, Eduard Anton, Thuy Oesterreich, Pascal Meier, and Frank Teuteberg. 2021. *Augmenting Humans in the Loop: Towards an Augmented Reality Object Labeling Application for Crowdsourcing Communities.* Springer, New York, NY, USA. https://www.springerprofessional.de/en/augmenting-humans-in-the-loop-towards-an-augmented-reality-objec/19762536
[83] SensorTips. 2022. *What sensors are used in AR/VR systems?* SensorTips. Retrieved May 2, 2023 from https://www.sensortips.com/featured/what-sensors-are-used-in-ar-vr-systems-faq/
[84] F. Shear. 2023. *Army accepts prototypes of the most advanced version of IVAS.* U.S. Army. https://www.army.mil/article/268702/army_accepts_prototypes_of_the_most_advanced_version_of_ivas
[85] Alan Smith, Charlie Duff, Rodrigo Sarlo, and Joseph L. Gabbard. 2022. Wearable Augmented Reality Interface Design for Bridge Inspection. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).* 497–501. https://doi.org/10.1109/VRW55335.2022.00111
[86] Riccardo Spezialetti, Darren Jiajun Tan, Alessandro Tonioni, Keisuke Tateno, and Federico Tombari. 2020. A Divide et Impera Approach for 3D Shape Reconstruction from Multiple Views. In *2020 International Conference on 3D Vision (3DV).* IEEE, 160–170. https://doi.org/10.1109/3DV50981.2020.00026

[87] Thilo Spinner, Udo Schlegel, Hanna Schäfer, and Mennatallah El-Assady. 2020. explAIner: A Visual Analytics Framework for Interactive and Explainable Machine Learning. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 1064–1074. https://doi.org/10.1109/TVCG.2019.2934629

[88] Wenjuan Sun, Paolo Bocchini, and Brian D. Davison. 2020. Applications of artificial intelligence for disaster management. *Natural Hazards* 103, 3 (01 Sep 2020), 2631–2689. https://doi.org/10.1007/s11069-020-04124-3

[89] Keisuke Tateno, Federico Tombari, Iro Laina, and Nassir Navab. 2017. CNN-SLAM: Real-Time Dense Monocular SLAM with Learned Depth Prediction. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 6565–6574. https://doi.org/10.1109/CVPR.2017.695

[90] Keisuke Tateno, Federico Tombari, and Nassir Navab. 2018. Deep object pose estimation for semantic robotic grasping of household objects. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 375–382.

[91] Michael Thees, Sebastian Kapp, Martin P. Strzys, Fabian Beil, Paul Lukowicz, and Jochen Kuhn. 2020. Effects of augmented reality on learning and cognitive load in university physics laboratory courses. *Computers in Human Behavior* 108 (2020), 106316. https://doi.org/10.1016/j.chb.2020.106316

[92] Marcus Valtonen Örnhag, Patrik Persson, Mårten Wadenbäck, Kalle Åström, and Anders Heyden. 2022. Trust Your IMU: Consequences of Ignoring the IMU Drift. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 4467–4476. https://doi.org/10.1109/CVPRW56347.2022.00493

[93] Jared Van Dam, Alexander Krasner, and Joseph L. Gabbard. 2020. Augmented Reality for Infrastructure Inspection with Semi-autonomous Aerial Systems: An Examination of User Performance, Workload, and System Trust. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 742–743. https://doi.org/10.1109/VRW50115.2020.00222

[94] Stylianos I. Venieris, Ioannis Panopoulos, and Iakovos S. Venieris. 2021. OODIn: An Optimised On-Device Inference Framework for Heterogeneous Mobile Devices. In *2021 IEEE International Conference on Smart Computing (SMARTCOMP)*. 1–8. https://doi.org/10.1109/SMARTCOMP52413.2021.00021

[95] Daniel Wagner, Thomas Pintaric, Florian Ledermann, and Dieter Schmalstieg. 2005. Towards Massively Multi-user Augmented Reality on Handheld Devices. In *Pervasive Computing*, Hans W. Gellersen, Roy Want, and Albrecht Schmidt (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 208–219.

[96] Johanna Wald, Keisuke Tateno, Jürgen Sturm, Nassir Navab, and Federico Tombari. 2018. Real-Time Fully Incremental Scene Understanding on Mobile Platforms. *IEEE Robotics and Automation Letters* 3, 4 (2018), 3402–3409. https://doi.org/10.1109/LRA.2018.2852782

[97] Runze Wang, Huimin Lu, Junhao Xiao, Yi Li, and Qihang Qiu. 2018. The Design of an Augmented Reality System for Urban Search and Rescue. In *2018 IEEE International Conference on Intelligence and Safety for Robotics (ISR)*. 267–272. https://doi.org/10.1109/IISR.2018.8535823

[98] Karl E. Weick. 1995. *Sensemaking in Organizations*. Sage Publications.

[99] Christopher D. Wickens, Justin G. Hollands, Simon Banbury, and Raja Parasuraman. 2013. *Engineering Psychology and Human Performance* (4 ed.). Psychology Press. https://doi.org/10.4324/9781315665177

[100] Chathurika S. Wickramasinghe, Daniel L. Marino, Javier Grandio, and Milos Manic. 2020. Trustworthy AI Development Guidelines for Human System Interaction. In *2020 13th International Conference on Human System Interaction (HSI)*. 130–136. https://doi.org/10.1109/HSI49210.2020.9142644

[101] Matthew Wilchek, Will Hanley, Jude Lim, Kurt Luther, and Feras A. Batarseh. 2023. Human-in-the-loop for computer vision assurance: A survey. *Engineering Applications of Artificial Intelligence* 123 (2023), 106376. https://doi.org/10.1016/j.engappai.2023.106376

[102] Fan Yang, Zhiwen Yu, Liming Chen, Jiaxi Gu, Qingyang Li, and Bin Guo. 2021. Human-Machine Cooperative Video Anomaly Detection. 4, CSCW3 (2021). https://doi.org/10.1145/3434183

[103] Ying Yang, Tim Dwyer, Michael Wybrow, Benjamin Lee, Maxime Cordeil, Mark Billinghurst, and Bruce H. Thomas. 2022. Towards Immersive Collaborative Sensemaking. 6, ISS, Article 588 (nov 2022), 25 pages. https://doi.org/10.1145/3567741

[104] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. 2018. Object Detection with Deep Learning: A Review. *CoRR* abs/1807.05511 (2018). arXiv:1807.05511 http://arxiv.org/abs/1807.05511