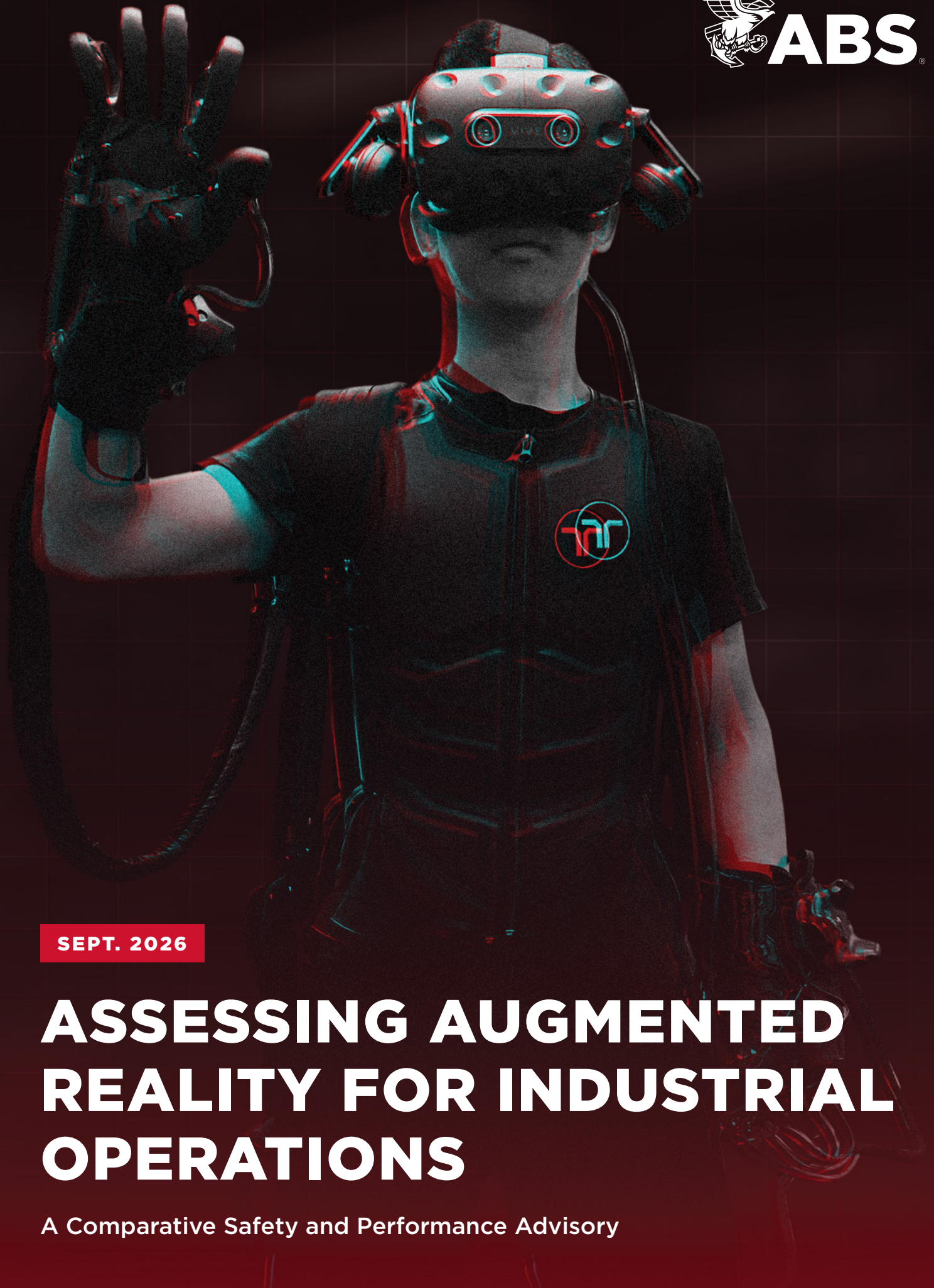




SEPT. 2026

ASSESSING AUGMENTED REALITY FOR INDUSTRIAL OPERATIONS

A Comparative Safety and Performance Advisory

Table of Contents

SECTION 01: INTRODUCTION 2

SECTION 02: AR IN INDUSTRIAL ENVIRONMENTS 4

2.1 AR within the Broader XR Landscape 4

2.2 Core Functional Capabilities of Augmented Reality 5

2.3 Representative Industrial Applications of Augmented Reality 6

2.4 Considerations Affecting Use of AR in Industrial Environments 8

2.5 Existing Standards, Regulatory Considerations and Related Frameworks 9

2.6 Relevance to Comparative Assessment 10

SECTION 03: SAFETY, SITUATIONAL AWARENESS AND HUMAN PERFORMANCE CONSIDERATIONS 12

3.1 Relationship between Usability, Human Performance and Safety 12

3.2 Potential Mechanisms for Safety and Performance Improvement 12

3.3 Potential Mechanisms for Safety and Performance Degradation 14

3.4 Situational Awareness 14

3.5 Task Dependence and Environmental Dependence 14

3.6 Technology-Induced Failure Modes and New Risk Pathways 16

3.7 Implications for Assessment 16

SECTION 04: METHODOLOGY FOR COMPARATIVE SAFETY AND PERFORMANCE ASSESSMENT 18

4.1 Basis of Comparison 18

4.2 Assessment Objectives 20

4.3 Stepwise Comparative Assessment Process 21

4.4 Comparative Assessment Measures 26

4.6 Evaluation Conditions and Evidence Quality 28

4.7 Assessment Outcomes and Suitability Determinations 30

4.8 Application of the Methodology 31

4.9 Summary 31

SECTION 05: IMPLEMENTATION CONSIDERATIONS AND RECOMMENDATIONS 32

5.1 Select Use Cases Deliberately 32

5.2 Define the Baseline Before Evaluating the Technology 34

5.3 Evaluate Safety and Performance Together 34

5.4 Validate Under Representative Conditions 34

5.5 Apply a Phased Deployment Approach 34

5.6 Address Human Factors Explicitly 36

5.8 Do Not Infer Operational Suitability from Training Value Alone 37

5.9 Monitor Performance and Residual Risk After Deployment 37

5.10 Recommendations 38

REFERENCES 40

APPENDIX: SELECTING MEASUREMENT APPROACHES FOR COMPARATIVE ASSESSMENT 42

Forward

Augmented reality (AR) is increasingly being considered for industrial applications such as inspection, maintenance, navigation, training and remote assistance. In these settings, AR can improve access to task-relevant information, support procedural consistency and enhance decision-making. At the same time, it may introduce new risks related to distraction, visual occlusion, cognitive workload, interface complexity, overreliance on system outputs and degraded situational awareness.

For this reason, the suitability of AR should not be judged on technical capability alone. Augmented reality should improve or at least maintain acceptable safety and operational performance when compared with the current method of work for the same task.

This advisory presents a structured methodology for comparatively assessing AR-enabled tasks against baseline work practices across two principal dimensions:

- Comparative safety, including situational awareness, workload, error potential and technology-induced hazards or failure modes
- Comparative performance, including task execution, accuracy, efficiency, quality and operational effectiveness

The intent is to provide a consistent basis for determining when AR may be suitable, when limitations or additional controls may be required, and when its use may not be appropriate for a given application. In this way, AR evaluation can be aligned with established engineering practice, human factors principles and operational safety objectives.



SECTION 01

Introduction

Augmented reality (AR) is a form of human-machine interaction in which digital information is presented in relation to the user’s physical environment during task execution. In industrial applications, AR may be delivered through head-mounted displays, handheld devices or other interfaces that provide task-relevant information at the point of use. Depending on the configuration, these systems may affect usability, field of view, attention management, physical burden and operational practicality in different ways.

Interest in AR for industrial operations has grown as digital technologies have become more capable and accessible. Commonly cited benefits include improved access to information, reduced reliance on memory, greater consistency in task execution, and enhanced support for training, guidance, and decision-making. In some applications, AR may reduce task time, improve procedural compliance or assist personnel in identifying relevant equipment conditions or hazards.

These benefits do not come automatically with the technology. The impact of AR depends on the task being performed, the user, the operating environment and the design of the interface itself. As shown in Table 1, AR systems may also introduce tradeoffs such as distraction, increased cognitive workload, altered visual scanning and decision patterns, new error pathways, and reduced reliability due to system limitations. These effects cannot be understood from system features alone. They must be evaluated relative to the existing method of work.

A meaningful evaluation requires comparison between:

- A **baseline condition**, representing the current accepted method of performing the task
- A **technology-enabled condition**, representing the same task performed using a defined AR application, device configuration and workflow

This comparison is necessary because the introduction of AR may affect both safety-related and performance-related outcomes. An AR application may improve access to information or procedural consistency while also increasing distraction, altering situational awareness, or introducing new dependencies on software, connectivity, or device performance.

For the purposes of this advisory, the assessment is structured across two principal dimensions:

- **Safety**, including situational awareness, workload, error potential and technology-induced hazards or failure modes
- **Performance**, including task completion, efficiency, accuracy, quality and operational effectiveness

This approach supports a more disciplined and defensible basis for decisions regarding pilot deployment, operational limitations, broader implementation or rejection of a given AR application.

The sections that follow examine representative industrial applications of AR, consider how its use may influence safety and human performance, and present a comparative methodology for evaluating suitability relative to existing work practices.

AR ASPECT	POTENTIAL BENEFIT	TRADEOFF/RISK INTRODUCED	IMPACT AREA	EXAMPLE IMPLICATION
Digital information overlay	Improved access to relevant information	Competes with visual attention and may obscure real-world elements	Situational awareness	Critical physical cues may be missed while focusing on overlays
Visual augmentation	Enhanced guidance and context	Alters normal visual scanning and decision-making patterns	Decision quality	Users may rely on AR cues instead of natural environment checks
Information density	Faster comprehension of complex data	Increased cognitive workload	Cognitive performance	Mental fatigue or slower responses under stress
Task guidance	Reduced task completion time	Introduction of new error pathways	Task accuracy	Automation bias leading to incorrect actions
System responsiveness	Real-time feedback and updates	Latency or inaccurate registration	Usability and trust	Delayed or misaligned overlays reduce confidence in the system
Connectivity	Continuous data exchange	Communication loss	Operational reliability	Sudden loss of AR support mid-task
Display quality	Hands-free visualization	Poor visibility in certain lighting or environments	Field usability	Difficult to read overlays outdoors or in bright conditions
Power supply	Portable, wearable operation	Battery depletion	Sustained operations	System becomes unavailable during critical tasks
Alerting mechanisms	Faster attention to important events	Poorly designed alerts cause distraction or overload	Human factors	Missed or ignored alerts due to alarm fatigue
Training applications	Improved training effectiveness	Unsuitable for safety critical field use	Training vs. operations	Effective for learning but risky in live environments
Overall system adoption	Enhanced information access	Increased distraction	Human performance	Net benefit depends on task context

Table 1: Key tradeoffs associated with the use of AR systems.



SECTION 02

AR in Industrial Environments: Functional Capabilities, Applications and Use Considerations

This section summarizes the characteristics of augmented reality most relevant to its use in industrial operations. Rather than offering a broad overview of extended reality (XR) technologies, it focuses on the features, use cases and operating conditions that affect whether AR is suitable for a given task.

In industrial settings, AR is best understood as a human-machine interface that can influence attention, workload, decision-making and situational awareness during work execution. Its evaluation should address more than the information the technology can present. It should also consider how its use changes task performance relative to existing practice.

2.1 AR WITHIN THE BROADER XR LANDSCAPE

Augmented reality is part of the broader class of XR technologies. For the purposes of this advisory, the distinctions matter only when they affect operational use as shown in Figure 1.

This distinction is important because industrial personnel using AR must continue to monitor equipment, surroundings, hazards, coworkers and changing operating conditions while performing work. Augmented reality should be assessed as an interface embedded in the task environment, not simply as an added information source.



Figure 1: VR, AR and MR are part of the broader XR spectrum; their distinctions are primarily significant when they affect operational use.

2.2 CORE FUNCTIONAL CAPABILITIES OF AUGMENTED REALITY

Augmented reality systems may support industrial activities through a range of functional capabilities. The specific implementation varies by device type, software configuration and intended use case, but commonly cited capabilities are shown in Figure 2.

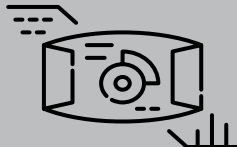
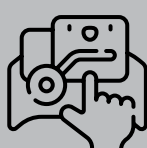
	VISUAL INFORMATION OVERLAY Presentation of digital information within or adjacent to the user's field of view, including text, icons, operating parameters, reference data and procedural prompts.
	CONTEXTUAL INFORMATION DELIVERY Display of information associated with a particular asset, location, operating state or task step.
	SPATIAL ANNOTATION AND OBJECT IDENTIFICATION Highlighting, labeling, or marking of equipment, components, pathways, boundaries, or inspection points.
	PROCEDURAL GUIDANCE Step-by-step work instructions, checklists, confirmation prompts, progression logic or task sequencing support.
	REMOTE COLLABORATION AND EXPERT ASSISTANCE Sharing of the user's perspective with a remote party, with the ability to provide verbal direction or visual annotation.
	ALERTING AND DECISION SUPPORT Notifications related to hazards, abnormal conditions, operational status or required actions.
	VISUALIZATION OF NON-VISIBLE OR ABSTRACT INFORMATION Presentation of data that is not otherwise directly observable, such as sensor outputs, historical information, asset status, design intent or digital model content.

Figure 2: Core functional capabilities of AR.

These capabilities may provide value by reducing the need to shift attention between the task and external information sources, reducing reliance on memory, improving procedural consistency or accelerating access to relevant information.

However, the presence of these capabilities does not establish operational benefits by itself. The same functions may also compete with visual attention, increase cognitive burden and alter normal visual scanning behavior. They can also create dependence on system accuracy, timing and reliability. Their value therefore depends on the task, the interface design and the conditions of use.

2.3 REPRESENTATIVE INDUSTRIAL APPLICATIONS OF AUGMENTED REALITY

Augmented reality may be applied across a range of industrial activities, including inspection, maintenance, navigation, training, remote assistance and hazard communication. The same underlying capability may have different effects depending on the use case.

Table 2 summarizes representative industrial applications of AR and highlights how the same underlying capabilities may introduce different benefits and risks depending on task context. Rather than being uniformly beneficial, AR affects safety and performance differently across use cases and must be evaluated relative to the baseline method of work.



APPLICATION AREA	PRIMARY AR FUNCTIONS	POTENTIAL BENEFITS	KEY RISKS/ CONCERNS	PRIMARY ASSESSMENT FOCUS
Procedural guidance and work instruction	Step-by-step overlays, prompts, confirmations	Improved consistency, reduced memory load, support for less experienced users	Visual distraction, timing issues, overfocus on display	Situational awareness, task interruption, error pathways
Inspection and condition assessment	Equipment data, inspection points, tolerances	Improved completeness, consistency, documentation	Obstructed view, altered visual scanning behavior, overconfidence	Observation quality, independent judgment
Navigation and location based awareness	Route guidance, hazard zones, wayfinding	Reduced search time, improved orientation	Latency, misalignment, attention diversion	Hazard awareness, reliability in dynamic environments
Remote assistance and expert support	Shared view, annotations, live guidance	Access to expertise, reduced delays, knowledge transfer	Communication dependence, divided attention	Workload, trust, communication reliability
Training and competency development	Guided workflows, simulated conditions	Improved engagement, retention, rehearsal	Training success does not equal field suitability	Transfer to operational use
Hazard notification and decision support	Alerts, exclusion zones, status indicators	Faster hazard recognition	Alert fatigue, overreliance, false cues	Trust calibration, error consequences
Display quality	Hands-free visualization	Poor visibility in certain lighting or environments	Field usability	Difficult to read overlays outdoors or in bright conditions
Power supply	Portable, wearable operation	Battery depletion	Sustained operations	System becomes unavailable during critical tasks
Alerting mechanisms	Faster attention to important events	Poorly designed alerts cause distraction or overload	Human factors	Missed or ignored alerts due to alarm fatigue
Training applications	Improved training effectiveness	Unsuitable for safety critical field use	Training vs. operations	Effective for learning but risky in live environments
Overall system adoption	Enhanced information access	Increased distraction	Human performance	Net benefit depends on task context

Table 2: Representative industrial applications of AR and key assessment considerations.

These examples illustrate that AR does not have uniform benefits. Suitability depends on whether the AR-enabled condition improves or at least maintains acceptable safety and task performance under the intended conditions of use.

2.4 CONSIDERATIONS AFFECTING USE OF AR IN INDUSTRIAL ENVIRONMENTS

The suitability of AR depends on both software functionality and the interactions between the task, the user, the interface and the operating environment. In industrial settings, these factors may materially affect both device performance and human performance.

Figure 3 summarizes these factors affecting AR performance and safety.

Industrial environments may also introduce constraints such as motion, vibration, changing

lighting, weather exposure, confined spaces, communication limitations and personal protective equipment (PPE) requirements. These conditions can materially affect both the usability of AR and the consequences of degraded performance.

Augmented reality should not be treated as a neutral layer added to existing work. In live use, it becomes part of the operational task environment and may change how personnel perceive, interpret and respond to conditions in real time.

2.5 EXISTING STANDARDS, REGULATORY CONSIDERATIONS AND RELATED FRAMEWORKS

The use of AR in industrial environments may be influenced by existing standards, regulations and guidance related to occupational safety, human factors, ergonomics, human-machine interface design, equipment safety, communications, cybersecurity and sector-specific operating controls. Key standards and practices are presented in Table 3.

STANDARD	TITLE	DESCRIPTION
ANSI/CAN/UL 8400	Virtual Reality, Augmented Reality and Mixed Reality Technology Equipment	Safety standards addressing hazards and safety requirements for VR/AR/MR technology equipment.
ISO/IEC 18023	Sensor representation in mixed and augmented reality	Framework for representing sensor-based information in mixed and augmented reality.
ISO/IEC 18039	Mixed and Augmented Reality	Framework for mixed and augmented reality applications, components, systems, services and specifications.
ISO/IEC 18040	Live actor and entity representation in mixed and augmented reality (MAR)	Reference model and components for representing and controlling mixed and augmented reality scenes.
ISO/IEC 18520	Benchmarking of vision-based spatial registration and tracking methods for mixed and augmented reality (MAR)	Methods for assessing spatial alignment and tracking performance in AR/MR systems.
ISO/IEC 23000	Multimedia application format (MPEG-A)	Container and application formats used for interoperable multimedia (including immersive media formats within the series).
ISO/IEC 5927	Augmented and virtual reality safety; Guidance on safe immersion, set up and usage	Guidance for safe setup and use of AR/VR devices, with a focus on workplace health and safety.
IEC 62368-1	Audio/video, information and communication technology equipment Part 1: Safety requirements	Hazard-based safety standard for AV/ICT equipment; often applicable to wearable and computing devices used in industrial contexts.
ISO 9241 (series)	Ergonomics of human-system interaction	Usability and human-centered design guidance relevant to AR interface design and safety-critical interaction.
IEEE P2048 (and related IEEE metaverse/XR efforts)	Metaverse terminology, definitions and taxonomy (draft)	Terminology work that may inform consistent XR language (not a safety standard).

Table 3: Selected standards and frameworks relevant to AR/XR use in industrial environments

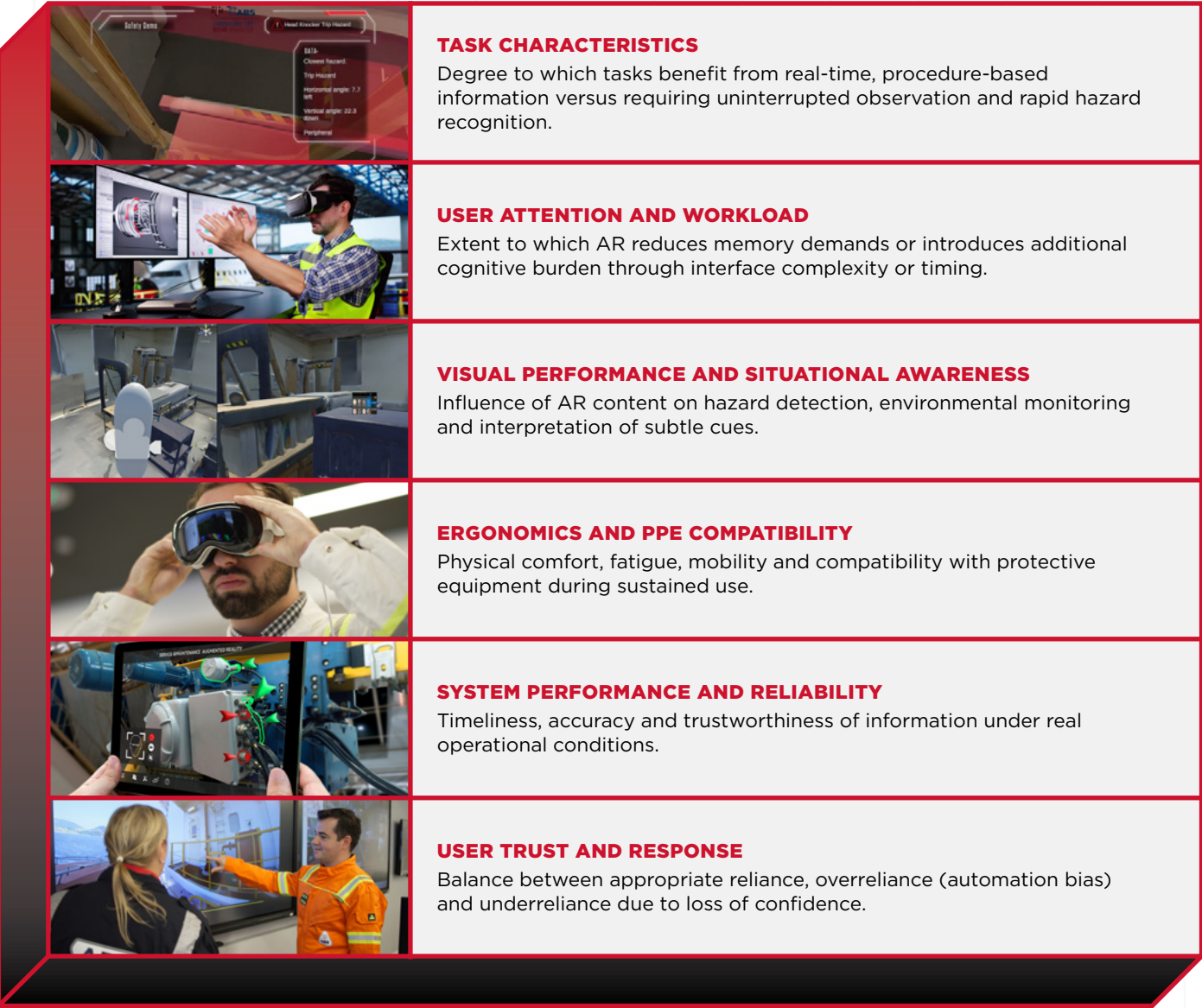


Figure 3: Considerations affecting AR performance and safety

Depending on the application and jurisdiction, relevant frameworks may include considerations associated with:

- Worker health and safety
- Human factors and ergonomics
- Usability of human-machine interfaces
- Alarm and alert design
- Machinery and equipment safety
- Compatibility with required PPE
- Electrical and electronic device suitability for the operating environment
- Wireless communications and cybersecurity
- Training, competence and supervision
- Management of change
- Validation, documentation and approval processes

In some cases, additional expectations may apply where AR is used in hazardous areas, safety-critical tasks, marine or offshore environments, regulated inspection activities, or other higher-consequence settings.

These frameworks are important, but they are not sufficient to establish operational suitability. Compliance with device, usability or general safety requirements does not by itself demonstrate that a specific AR application is appropriate for a specific task. Suitability still depends on how the AR-enabled condition affects safety and performance relative to the current method of work.

2.6 RELEVANCE TO COMPARATIVE ASSESSMENT

Augmented reality in industrial operations should not be evaluated solely based on technical capability, general usability or standards of conformance. The relevant question is whether the technology improves or at least maintains acceptable outcomes when compared with the baseline method for the same task.

This requires comparative assessment because:

- Users must continue to perceive and respond to the real environment while using AR
- Augmented reality may improve access to information while also introducing new attentional, interface and reliability risks
- The effects of AR depend strongly on task type, user characteristics and operating conditions
- Compliance with general standards does not by itself demonstrate task-level suitability

For these reasons, AR should be assessed against the current accepted method of work across both safety and performance outcomes. The following section examines these mechanisms in greater detail by considering how AR may influence safety, situational awareness and human performance during industrial task execution.



SECTION 03

Safety, Situational Awareness and Human Performance Considerations

The use of augmented reality in industrial tasks can affect access to information as well as how personnel allocate attention, interpret conditions, make decisions and perform work in real time. These effects may improve or degrade task execution depending on the demands of the task, the interface design, the operating environment and the reliability of the system.

For this reason, AR suitability should be evaluated based on its effects on human performance and safety during actual task execution, not functionality alone.

3.1 RELATIONSHIP BETWEEN USABILITY, HUMAN PERFORMANCE AND SAFETY

Usability in AR is more than a convenience issue. In industrial settings, the design of the interface may directly influence workload, attention, situational awareness, communication and error likelihood.

This relationship may be understood as progressive: interface design and usability influence workload, attention, and human performance, which in turn affect situational awareness and safety outcomes.

The evaluation of AR should consider whether the technology provides useful information and whether the way in which that information is delivered supports or degrades effective task behavior under operational conditions.

3.2 POTENTIAL MECHANISMS FOR SAFETY AND PERFORMANCE IMPROVEMENT

Augmented reality may improve safety or task performance where it provides relevant information at the point of use without materially degrading awareness of the surrounding environment. Potential mechanisms for improvement are shown in Table 4.



AR MECHANISM	DESCRIPTION	PRIMARY TASK BENEFIT	TYPICAL CONDITIONS WHERE MOST APPLICABLE
Improved access to task relevant information	Presentation of procedures, parameters or reference data at the point of task execution	Reduced interruption; improved continuity and procedural adherence [Gerdenitsch et al, 2022; Marques et al, 2025].	Information intensive or procedure driven tasks
Reduced reliance on memory	Step confirmation, condition checks and equipment identification provided in context	Lower cognitive burden [Moncur et al, 2023]; reduced omission risk (Bellucci et al, 2018)	Complex, multi step or infrequent tasks
Improved procedural consistency	Stepwise prompts, checklists and contextual guidance	Standardized execution; improved compliance [Liberty et al, 2026]	Tasks with quality or safety critical procedural requirements
Enhanced orientation and condition awareness	Identification of equipment, inspection points, paths, boundaries or abnormal conditions	Faster recognition; improved field decision-making [Moncur et al, 2023]	Complex layouts or visually dense environments
Support for training and less experienced personnel	Contextual guidance and in situ learning support	Improved learning, familiarization, and confidence [Han et al, 2022; Miranda et al, 2025]	Onboarding, refresher training or low experience users
Hazard notification and decision support	Alerts, exclusion zones, status indicators	Faster hazard recognition	Alert fatigue, overreliance, false cues
Display quality	Hands-free visualization	Poor visibility in certain lighting or environments	Field usability
Power supply	Portable, wearable operation	Battery depletion	Sustained operations
Alerting mechanisms	Faster attention to important events	Poorly designed alerts cause distraction or overload	Human factors
Training applications	Improved training effectiveness	Unsuitable for safety critical field use	Training vs. operations
Overall system adoption	Enhanced information access	Increased distraction	Human performance

Table 4: Potential mechanisms by which AR may support safety and situational awareness.

These benefits may be meaningful, particularly for information-intensive or procedure-driven work. However, they should not be assumed to be. Their value depends on the information being accurate, timely, appropriately prioritized and compatible with the task and environment.

3.3 POTENTIAL MECHANISMS FOR SAFETY AND PERFORMANCE DEGRADATION

The same characteristics that make AR useful may also create new sources of degraded performance or increased risk. Table 5 discusses potential mechanisms by which AR may degrade safety and situational awareness.

These mechanisms are central to suitability assessment because they may not appear in a feature-based description of the technology, but they can materially affect task safety and performance.

3.4 SITUATIONAL AWARENESS

Situational awareness is a critical consideration in the use of AR for live industrial work. In general terms, situational awareness involves perception of relevant elements in the environment, comprehension of their significance and anticipation of how conditions may change.

Augmented reality may support situational awareness when it helps users detect, interpret or prioritize relevant information without diverting attention from the physical environment. For example, AR may assist with identifying equipment status, boundaries, inspection locations or procedural context.

It may also degrade situational awareness if digital content competes with environmental observation, narrows attention, interrupts visual scanning behavior or encourages undue concentration on the display. In such cases, information availability may improve while awareness of the broader operating context degrades.

This issue is especially important in dynamic or hazard-rich environments where personnel must continuously monitor footing, weather, moving equipment, changing conditions, nearby activity or other operational cues. In such settings, even useful information may become counterproductive if its presentation interferes with continued awareness of the real environment.

The effect of AR on situational awareness depends less on the volume of information provided than on whether that information is presented in a manner compatible with safe task execution.

3.5 TASK DEPENDENCE AND ENVIRONMENTAL DEPENDENCE

The effects of AR are strongly dependent on both task type and operating conditions. A given AR application may be suitable where work is structured, repetitive and information-intensive, yet unsuitable where work is dynamic, exposed, congested or highly dependent on uninterrupted environmental monitoring.

Important contextual influences may include:

- Complexity and variability of the task
- Time pressure and pace of work
- Need for continuous visual scanning
- Proximity to moving equipment or hazardous energy
- Environmental lighting and visibility
- Noise, vibration or motion
- Weather exposure
- Communications reliability
- Use of gloves, helmets, face shields or other personal protective equipment
- User experience and familiarity with both the task and the technology

These factors may materially change both the value and the risk of AR. Suitability should therefore be assessed under conditions representative of intended use rather than inferred from controlled demonstrations or generalized claims of benefit.

POTENTIAL DEGRADATION MECHANISM	DESCRIPTION	PRIMARY IMPACT ON TASK EXECUTION	TYPICAL CONDITIONS OF CONCERN
Divided attention and distraction [Cao et al., 2025; Khan et al., 2025]	Concurrent processing of digital content and physical surroundings may divert attention from equipment, coworkers or emerging hazards	Delayed hazard detection; reduced responsiveness	Dynamic, congested or hazard rich environments
Increased cognitive workload [Moncur et al., 2023]	Excessive, poorly organized, mistimed or difficult to navigate information may increase mental effort rather than reduce it	Slower decision-making; pacing errors; increased error likelihood	High time pressure; multitasking; frequent condition changes
Visual competition and occlusion [Wang et al., 2025; Le & Park, 2023]	Digital overlays may compete with or obscure physical cues, alter visual scanning behavior or interfere with depth perception	Missed cues; degraded visual performance	Limited field of view; bright or cluttered displays; frequent refocusing
Automation bias and overreliance [Parasuraman & Riley, 1997]	Users may place excessive trust in AR provided guidance or alerts and reduce independent verification	Acceptance of incorrect prompts; missed discrepancies	Decision support or alert heavy applications
Disruption of established work practices [Bailey et al., 2022; Ito et al., 2021; Juliano et al., 2022]	Changes in task pacing, sequence, posture or coordination may interfere with practiced routines	Reduced efficiency; increased collision or coordination risk	Use of head mounted devices; unfamiliar interfaces
Degraded trust and inconsistent use [Howard et al., 2023]	Latency, false alerts, registration errors or awkward controls may reduce confidence and lead to intermittent reliance	Unpredictable task behavior; inconsistent procedure execution	Systems with variable reliability or immature design
Display quality	Hands-free visualization	Poor visibility in certain lighting or environments	Field usability
Power supply	Portable, wearable operation	Battery depletion	Sustained operations
Alerting mechanisms	Faster attention to important events	Poorly designed alerts cause distraction or overload	Human factors
Training applications	Improved training effectiveness	Unsuitable for safety critical field use	Training vs. operations
Overall system adoption	Enhanced information access	Increased distraction	Human performance

Table 5: Potential mechanisms by which AR may degrade safety and performance.

3.6 TECHNOLOGY-INDUCED FAILURE MODES AND NEW RISK PATHWAYS

The introduction of AR may create failure modes that are absent, less significant, or differently managed in the baseline method of work. These should be treated as part of the operational risk profile of the AR-enabled condition. Examples are included in Table 6.

These failure modes do not necessarily preclude the use of AR, but they demonstrate that AR can introduce new pathways to degraded performance or unsafe outcomes. They should therefore be considered alongside traditional task hazards when evaluating suitability.

3.7 IMPLICATIONS FOR ASSESSMENT

Taken together, these considerations show that the effects of AR on safety and performance are context-dependent and should be evaluated comparatively rather than assumed from functionality alone.

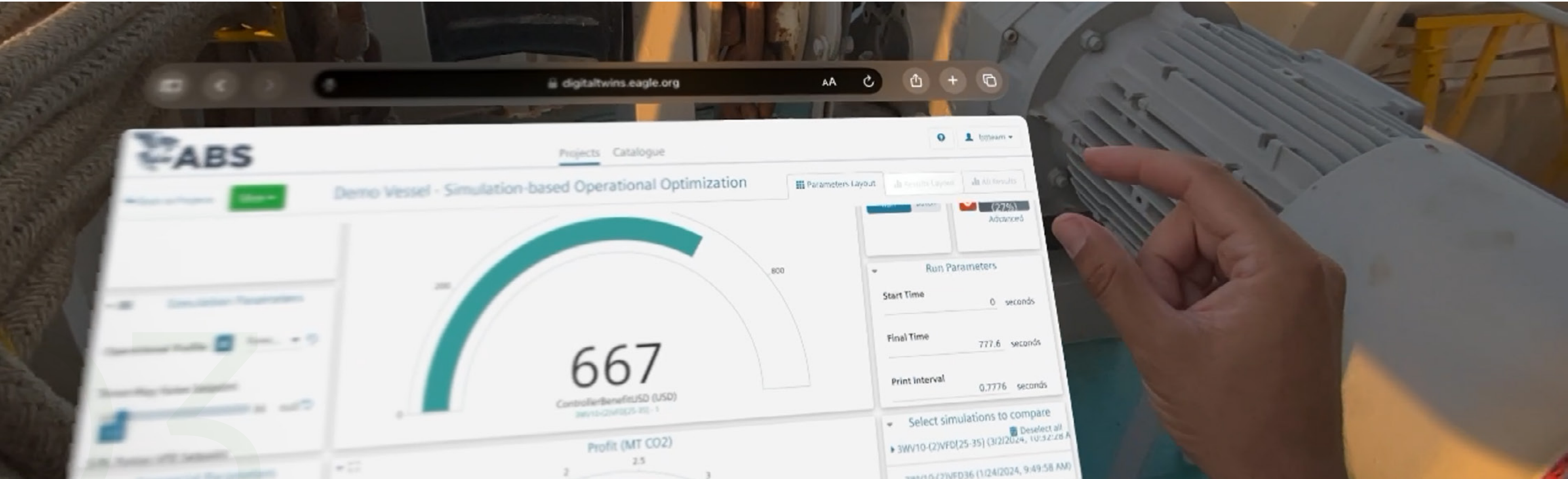
A meaningful assessment should consider how AR affects:

- Attention and workload
- Situational awareness
- Perception of hazards and abnormal conditions
- Task accuracy and consistency
- Error likelihood and recovery from interruption
- Efficiency and task completion
- Dependence on system reliability
- User trust and response behavior

The following section presents a structured methodology for evaluating these effects by comparing the AR-enabled condition with the baseline method of work for the same task.

FAILURE MODE	EXAMPLE CONSEQUENCE
Registration error	Misplaced guidance or misidentified equipment
Software errors	Unanticipated behavior or outcomes
Outdated information	Incorrect displays or recommendations
Latency	Delayed display of relevant conditions
Display clutter	Missed cues or reduced attention to surroundings
Connectivity loss	Loss of remote assistance or data support
Battery depletion	Mid-task system unavailability
False/missed alerts	Inappropriate response or missed hazard
Environment incompatibility	Ability to use device may be limited or degraded
User attention	Reduced monitoring due to overreliance on system outputs
Overall system adoption	Enhanced information access

Table 6: Examples of potential AR failure modes.



SECTION 04

Methodology for Comparative Safety and Performance Assessment

The use of augmented reality in industrial operations should be evaluated through comparison with the current accepted method of performing the same task. The purpose of this comparison is to determine whether the technology results in an acceptable or improved outcome relative to baseline practice, not simply that it functions as intended.

This advisory proposes a comparative assessment methodology based on two principal dimensions:

- **Comparative safety**, including effects on situational awareness, workload, error potential and exposure to new hazards or failure modes
- **Comparative performance**, including effects on task execution, efficiency, accuracy, consistency and operational effectiveness

The methodology is intended to support determinations regarding whether a given AR application is:

- Suitable for intended use
- Suitable with defined limitations or controls
- Suitable only for training or bounded use
- Not suitable for the intended application

4.1 BASIS OF COMPARISON

A meaningful assessment requires comparison between two defined conditions:

- **Baseline condition**, representing the current accepted method of performing the task
- **Technology-enabled condition**, representing the same task performed using a defined AR application, device configuration and associated workflow

The baseline condition should reflect actual current practice rather than an idealized or outdated procedure that is not representative of how the task is normally performed. Depending on the task, the baseline may include paper-based procedures, handheld devices, fixed displays, verbal coordination, conventional inspection methods or combinations of these.

The AR-enabled condition should be defined with sufficient clarity to support meaningful comparison. As applicable, this should include:

- Device type and display configuration
- Software version or functional configuration
- Information presented to the user
- Alerting or decision-support functions
- Connectivity requirements
- Remote-support functions
- Expected user interactions
- Intended operating conditions and limitations

Without a clearly defined baseline and AR-enabled condition, the assessment may become a general discussion of technology attributes rather than a defensible comparison of task outcomes.



4.2 ASSESSMENT OBJECTIVES

The comparative assessment should determine how the use of AR changes the performance of the task and the risks associated with it. The assessment should be structured to answer the following questions, as shown in Table 7.

These questions recognize that suitability is not determined by functionality alone. A system may provide useful features but still be unsuitable if it degrades situational awareness, increases workload beyond acceptable limits or creates unacceptable reliance on device performance or communications.

ASSESSMENT OBJECTIVE	KEY QUESTION ADDRESSED
Comparative safety	Does AR maintain or improve safety relative to baseline practice?
Comparative performance	Does AR maintain or improve task performance relative to baseline practice?
New risks and burdens	Does AR introduce new failure modes, dependencies or workload?
Conditions for use	Under what limitations, controls, or contexts is AR suitable?

Table 7: Key assessment objectives for AR comparative analysis.



4.3 STEPWISE COMPARATIVE ASSESSMENT PROCESS

A structured comparative process should be used to evaluate the operational suitability of AR applications. The process should include the steps outlined in Figure 4.

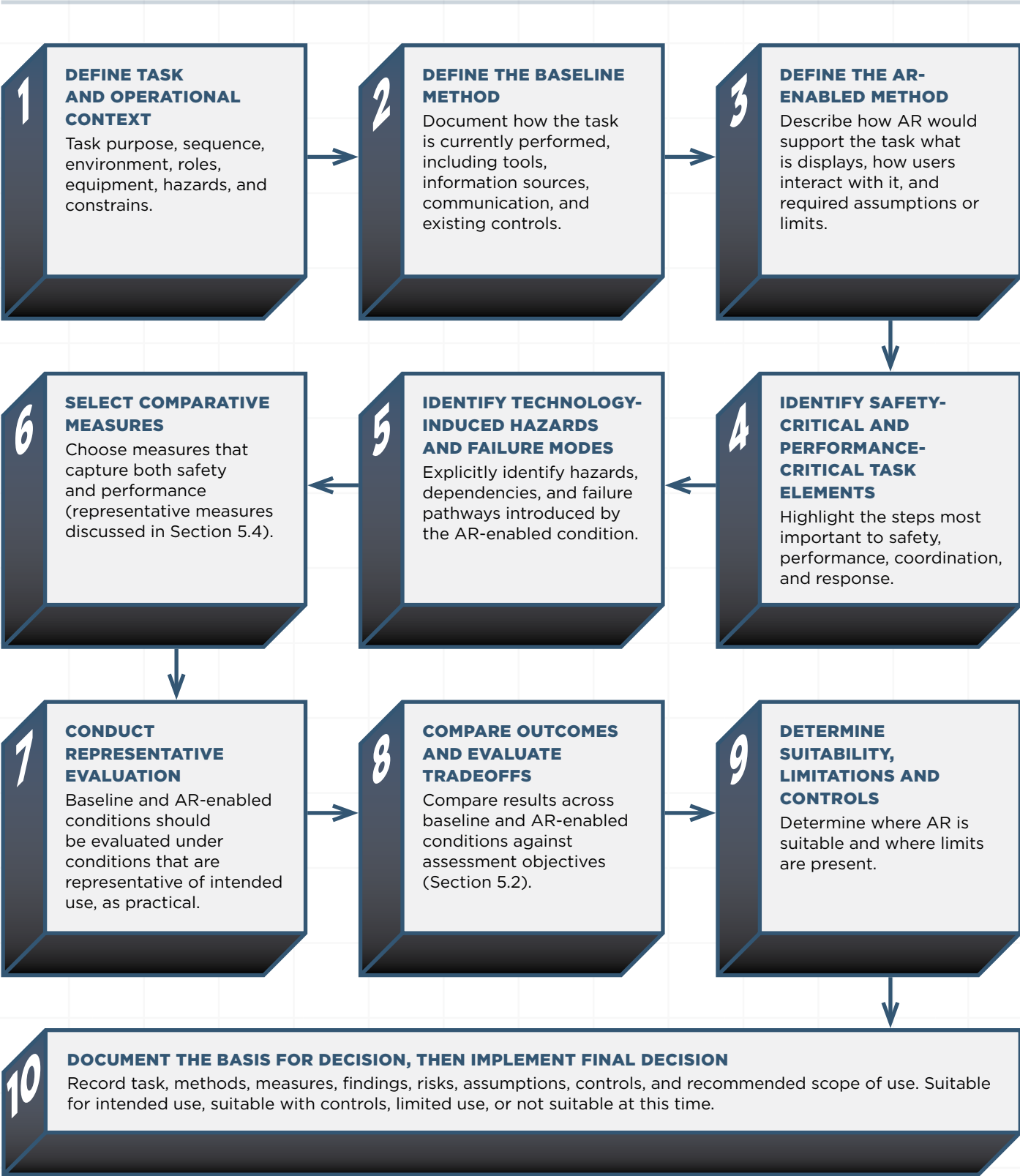


Figure 4: Stepwise Process for Evaluating AR Suitability in Operational Tasks.

Each step should be tailored to the task and operating context. Table 8 provides representative considerations that may be addressed at each step of the assessment.

STEP	EXAMPLE OF PRIMARY FOCUS OR CONSIDERATION
Step 1: Define the task and operational context	The task to be assessed should be defined in operational terms, including: • Task purpose and expected outcome • Task sequence and critical steps • Operating environment • Personnel roles and interactions • Equipment involved • Normal hazards and constraints • Time sensitivity or operational pressure • Relevant environmental conditions • Safety-critical aspects of execution This definition is necessary because the effect of AR depends on the context in which it is used. A technology that is suitable for one task or setting may not be suitable for another.
Step 2: Define the baseline method	The current accepted method of performing the task should be described in sufficient detail to serve as the reference condition. This should include: • Information sources used by personnel • Communication methods • Tools and interfaces currently used • Reliance on memory, procedures or expert support • Known limitations or error-prone aspects of the baseline approach • Current controls used to maintain safety and performance This step is important because AR should be compared against actual practice, not against an artificially weak baseline.
Step 3: Define the AR-enabled method	The proposed AR application should then be described. This description should include: • What information is displayed • When and how information is presented • How the user interacts with the system • Whether the system provides prompts, warnings, annotations or decision support • Whether remote assistance is involved • What infrastructure or connectivity is required • What assumptions are made regarding training, calibration, setup or device availability This step should also identify intended limitations of use, such as restricted environments, non-safety-critical applications or use only by trained personnel.

Table 8: Representative considerations at each step of the process (continued on following page).

STEP	EXAMPLE OF PRIMARY FOCUS OR CONSIDERATION
Step 4: Identify safety-critical and performance-critical task elements	The assessment should identify the elements of the task most relevant to safety and performance, such as: • Steps where omission or error could create harm • Points requiring careful observation or interpretation • Activities requiring coordination with others • Situations requiring rapid response or adaptation • Components of the task that drive duration, quality or consistency This helps focus on the evaluation of the areas where the technology is most likely to create meaningful benefits or meaningful degradation.
Step 5: Identify technology-induced hazards and failure modes	The assessment should explicitly identify hazards, dependencies, and failure pathways introduced by the AR-enabled condition. Examples may include: • Visual obstruction or display competition • Increased cognitive workload • Registration error or content misalignment • Latency or delayed updates • False or missed alerts • Connectivity loss • Power loss or device shutdown • Software error or outdated information • Overreliance on prompts or automated outputs • Incompatibility with personal protective equipment (PPE), weather, motion or lighting conditions These issues should be considered as part of the operational risk profile of the AR-enabled task, not as separate technical concerns.
Step 6: Select representative task basis and comparative measures	Comparative measures should be selected to evaluate the effect of AR on both safety and performance. The selected measures should be appropriate to the task and sufficiently sensitive to reveal meaningful differences between the baseline and AR-enabled conditions. Where direct evaluation of the intended industrial task is impractical, identify a representative metatask or task construct that captures the dominant human-performance demands relevant to the intended use case, and select comparative measures accordingly. Representative measures are discussed in 4.4.
Step 7: Conduct representative evaluation	The baseline and AR-enabled conditions should be evaluated under conditions that are representative of intended use to the extent practicable. Depending on the application, this may include simulation, pilot testing, controlled trials, field observation or staged operational evaluation. The evaluation should account for: • Realistic task conditions • Relevant environmental influences • Expected user population and experience level • Likely interruptions, communications demands and workflow conditions • Foreseeable deviations, failure conditions or abnormal scenarios Assessment based solely on idealized demonstrations may not provide a sufficient basis for determining suitability in operational use.

Table 8: Representative considerations at each step of the process (continued on following page).

STEP	EXAMPLE OF PRIMARY FOCUS OR CONSIDERATION
Step 8: Compare outcomes and evaluate tradeoffs	Results should be compared across the baseline and AR-enabled conditions to determine whether the technology: • Improves safety and performance • Improves performance without degrading safety • Improves some outcomes while degrading others • Introduces residual risks that require controls or limitations • Fails to provide sufficient benefit to justify added complexity This comparison should be evidence-based and should explicitly consider whether any claimed gains are offset by new burdens, dependencies or error pathways.
Step 9: Determine suitability, limitations and controls	Based on the comparison, the assessment should determine whether the AR application is: • Suitable for the intended operational use • Suitable with defined controls or limitations • Suitable only for selected tasks, environments or user groups • Suitable for training or rehearsal but not for live task execution • Not suitable for the intended use pending redesign, further validation or additional safeguards Where controls are required, these may include training, restricted use conditions, fallback procedures, supervision requirements, alert redesign, information simplification, connectivity requirements or operational exclusion from safety-critical tasks.
Step 10: Final decision	The basis for the suitability determination should be documented clearly. The record should describe: • The task assessed • The baseline and AR-enabled conditions • Measures used • Evaluation conditions • Key findings • Residual risks • Limitations and assumptions • Required controls • Recommended scope of use This documentation supports transparency, repeatability and management of future changes to the application or its operating context.

Table 8: Representative considerations at each step of the process.

Having established a structured process for defining the task, baseline and AR-enabled conditions, and identifying relevant hazards and tradeoffs, the next step is to determine how differences in safety and performance should be evaluated. The following section describes representative comparative measures that may be used to assess the effects of AR relative to baseline practice.



4.4 COMPARATIVE ASSESSMENT MEASURES

Comparative measures should be selected to reflect the actual purpose of the AR application and the aspects of task execution most relevant to the decision. In general, the assessment should include both safety and performance measures.

4.4.1 Task Demands and Representative Metatasks

The selection of comparative measures should reflect the nature of the task and the intended purpose of the AR application. Where the intended industrial task can be evaluated directly, measures should be selected against that task under representative conditions. Where direct replication is impractical due to cost, safety, access, operational disruption or task variability, the assessment may use a representative metatask.

A metatask is a generalized task construct that captures the dominant human-performance demands relevant to the intended application, even where the full industrial task is not reproduced directly. Representative metatask categories may include navigation, memory-loading, visual search and information processing, manual coordination, decision-making, interruption management, and supervisory control.

Representative metatask categories may include:

- **Navigation and wayfinding**, where the primary demands involve movement, route selection, orientation and environmental monitoring
- **Memory-loading tasks**, where the primary demands involve recall of steps, criteria, specifications or task status
- **Visual search and information processing tasks**, where the primary demands involve detection, discrimination, interpretation or monitoring of visual cues

- **Manual interaction or coordination tasks**, where the primary demands involve hand-eye coordination, manipulation, confirmation or sequencing
- **Decision-making tasks**, where the primary demands involve interpretation, prioritization, diagnosis or response selection
- **Interruption management or executive control tasks**, where the primary demands involve task switching, resumption, prioritization or supervisory control
- **Perceptual or sensory integration tasks**, where the primary demands involve combining visual, auditory, haptic or contextual cues during task execution

Metatasks may be useful for concept screening, prototype comparison, and early-stage evaluation of how AR affects workload, situational awareness, task behavior, error likelihood, or performance. However, they do not replace task-specific evaluations where direct assessment is feasible or necessary. Where metatasks are used, the basis for their representativeness and the limitations on inference should be documented explicitly, particularly for higher-consequence applications.

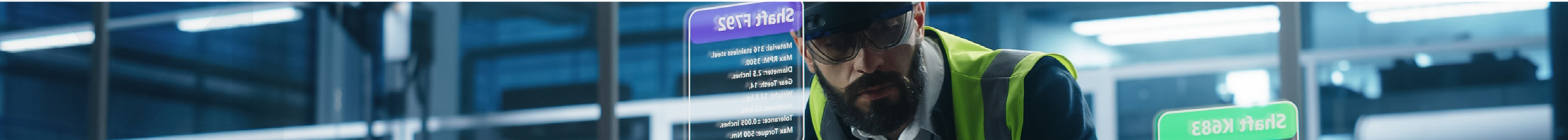
4.4.2 Representative Comparative Safety Measures

Representative comparative safety measures are summarized in Table 9. These measures help determine whether AR changes the user’s exposure to hazards, supports or degrades awareness, alters mental demand or creates new dependencies that affect safe task execution.

Additional discussion of measurement approach selection, including black-box, gray-box, and white-box methods, is provided in the Appendix.

ASSESSMENT AREA	REPRESENTATIVE MEASURES	EXAMPLE INTERPRETATION/USE
Hazard exposure and safety margin	Proximity to hazards; time to hazard or conflict; hazardous contact or near contact	Used to assess whether AR changes the user’s exposure to hazards relative to the baseline method
Situational awareness and attention	Hazard detection; ability to detect abnormal conditions; user attention and visual scanning behavior; situational awareness rating	Used to assess whether AR supports or degrades awareness of the surrounding environment during task execution
Workload and cognitive demand	Subjective workload rating; cognitive workload indicators; task interruption frequency; response to alarms, prompts, or warnings	Used to assess whether AR reduces or increases mental demand and attentional burden
Task error and procedural reliability	Procedural omission or commission errors; error likelihood at critical task steps; task recovery following interruptions	Used to assess whether AR improves execution reliability or creates new error pathways
Coordination and communication	Quality of coordination with other personnel; communication clarity; response timing in coordinated tasks	Used where task safety depends on shared awareness or interaction with others
Technology-induced dependence and failure exposure	Reliance on connectivity, power, tracking accuracy, or system performance; frequency of degraded-function events	Used to assess whether the AR-enabled condition introduces operational dependencies or new failure vulnerabilities not present in baseline practice
Trust and user response	Trust calibration; evidence of overreliance or under reliance; response to system prompts, warnings or recommendations	Used to assess whether users interact with the AR system in a manner consistent with safe task execution
Power supply	Portable, wearable operation	Battery depletion
Alerting mechanisms	Faster attention to important events	Poorly designed alerts cause distraction or overload
Training applications	Improved training effectiveness	Unsuitable for safety critical field use
Overall system adoption	Enhanced information access	Increased distraction

Table 9: Representative comparative safety measures.



4.4.4 Use of Leading and Lagging Measures

- Comparative assessment may involve both leading measures and lagging measures.
- **Leading measures** help identify developing issues before adverse outcomes occur. In AR assessment, these may include workload, attention allocation, display usability, trust calibration, visual scanning behavior and frequency of degraded-function events.
 - **Lagging measures** reflect realized outcomes, such as task errors, near misses, rework, task completion time, quality deficiencies or observed safety events.

Where practicable, both should be used. Leading measures may reveal emerging risks early but should not substitute for outcome-based evidence where meaningful task results can be observed.

Qualitative and quantitative indicators may also be used together, particularly where situational awareness, workload or trust cannot be fully captured through simple task metrics alone.

4.6 EVALUATION CONDITIONS AND EVIDENCE QUALITY

The value of a comparative assessment depends on the measures selected and the realism and rigor of the evaluation conditions. Evidence should be interpreted according to how well the assessment reflects intended operational use.

Stronger assessment bases are provided by evaluations described in Figure 5.

By contrast, limited demonstrations, vendor-led trials or assessments conducted only under highly controlled or idealized conditions may be useful for preliminary screening but may not be sufficient to support operational deployment decisions. Where metatasks are used in place of direct task replication, the representativeness of the metatask and the resulting limitations on inference should be documented explicitly.

Where evidence is limited, this should be stated explicitly, and conclusions regarding suitability should be correspondingly bound.



Figure 5: Comparative assessment evaluation considerations.



4.7 ASSESSMENT OUTCOMES AND SUITABILITY DETERMINATIONS

The comparative assessment should support a clear determination regarding the suitability of the AR application for the intended task and operating context. Results should be interpreted in terms of whether the AR-enabled condition provides an acceptable or improved outcome relative to the baseline method of work across both safety and performance dimensions. Different patterns of results may support different decisions.

Representative suitability determinations include the following:

- **Suitable for intended use**
Appropriate where the AR-enabled condition maintains or improves safety relative to baseline practice and provides meaningful operational benefit under the intended conditions of use.
- **Suitable with defined limitations or controls**
Appropriate where the AR application provides value, but only if specific limitations, restrictions or controls are applied.

- **Suitable for training or bounded use only**
Appropriate where the AR application supports training, rehearsal or selected low-consequence tasks, but does not provide a sufficient basis for broader live operational use.
- **Not suitable for intended use**
Appropriate where the AR-enabled condition introduces unacceptable degradation in safety, situational awareness, workload, reliability, task performance or where the benefits do not justify the burdens introduced.

Suitability conclusions should be framed in relation to the specific task, operating conditions, assumptions and limitations evaluated.

4.8 APPLICATION OF THE METHODOLOGY

This methodology may be applied at different stages of technology maturity, including concept screening, prototype evaluation, pilot deployment and pre-operational validation. In each case, the objective remains the same: to determine whether the AR-enabled condition produces an acceptable or improved outcome relative to the baseline method of work for the intended task.

The rigor of the assessment should be proportionate to:

- The significance of the task
- The degree of hazard exposure
- The consequence of error
- The extent to which AR changes task execution

For low-consequence or bounded applications, an initial comparative assessment may be sufficient to support limited use. For applications affecting safety-critical tasks, inspection quality, operational decision-making or worker hazard exposure, a more rigorous evaluation may be required, including representative operational testing and explicit consideration of degraded or failure conditions.

4.9 SUMMARY

A structured comparative methodology is necessary to determine whether an AR-enabled task provides an acceptable or improved result when compared with the current method of work. Such a methodology should define the baseline and AR-enabled conditions, identify critical task elements and new failure pathways, evaluate both safety and performance, and support a clear determination regarding suitability, limitations, and required controls.

This approach helps shift the evaluation of AR from general claims of technical capability to a more disciplined assessment of operational value and safety impact. It also provides a more consistent basis for determining where AR may be beneficial, where caution is warranted, and where the technology is not appropriate for the intended use.



SECTION 05

Implementation Considerations and Recommendations

Implementation decisions regarding augmented reality should not rely solely on demonstrated technical capability. They should rely on whether the AR-enabled condition provides acceptable safety and operational performance relative to baseline practice under the intended conditions of use.

The following considerations and recommendations are intended to support the practical application of the comparative assessment methodology described in Section 4.

5.1 SELECT USE CASES DELIBERATELY

Augmented reality should be applied selectively rather than broadly. The most suitable initial use cases are typically those in which:

- The task is information-intensive or procedure-driven
- Access to timely contextual information is likely to improve execution

- The baseline method has identifiable limitations
- The operational environment is sufficiently stable to support reliable use
- The consequences of distraction or degraded situational awareness are manageable

Greater caution is warranted where tasks are highly dynamic, require uninterrupted visual scanning, involve significant hazard exposure or depend on rapid adaptation to changing conditions. In such cases, AR may introduce additional burden or reduce awareness even where useful information is provided.

Use cases should therefore be selected based on a credible comparative case for improvement, not simply because AR is technically applicable. An example of metatasks and potential extended reality features which may aid in performance are provided in Table 11.

METATASK CATEGORY	DOMINANT DEMAND	EXAMPLE AR ASSESSMENT FOCUS
Navigation/wayfinding	Orientation, route following, environmental monitoring	Attention diversion, wayfinding efficiency, hazard awareness
Memory-loading	Step recall, specification recall, task-state memory	Workload reduction, omission errors, procedural consistency
Visual search/information processing	Cue detection, discrimination, monitoring	Visual scanning behavior, missed cues, speed/accuracy tradeoff
Manual coordination	Hand-eye coordination, sequencing, confirmation	Interaction burden, task precision, interruption effects
Decision-making	Interpretation, prioritization, response selection	Decision quality, overreliance, time to response
Executive control/interruption management	Task switching, resumption, supervisory control	Workload, resumption lag, prompt timing, error recovery
Trust and user response	Trust calibration; evidence of overreliance or under reliance; response to system prompts, warnings or recommendations	Used to assess whether users interact with the AR system in a manner consistent with safe task execution
Power supply	Portable, wearable operation	Battery depletion
Alerting mechanisms	Faster attention to important events	Poorly designed alerts cause distraction or overload
Training applications	Improved training effectiveness	Unsuitable for safety critical field use
Overall system adoption	Enhanced information access	Increased distraction

Table 11: Examples of task categories and potential XR features that may aid performance.



5.2 DEFINE THE BASELINE BEFORE EVALUATING THE TECHNOLOGY

A recurring implementation error is to evaluate AR against a poorly defined or unrealistic baseline. Before assessing the technology, the current method of work should be established clearly, including how information is presently accessed, how tasks are coordinated, where errors or inefficiencies occur and what existing controls support safe execution.

This is necessary both to avoid overstating AR benefits and to identify what aspects of current task execution must be preserved. If the baseline is not defined clearly, the assessment may become a demonstration of features rather than an evaluation of operational suitability.

5.3 EVALUATE SAFETY AND PERFORMANCE TOGETHER

Implementation decisions should not be based on performance gains alone. Reduced task time, improved information access or greater procedural support does not by itself justify operational use if AR degrades situational awareness, increases workload beyond acceptable limits or introduces unmanaged failure pathways.

Similarly, an application that appears acceptable from a narrow safety perspective may still provide insufficient operational value to justify its burden of deployment, training, maintenance and support.

Implementation decisions should therefore consider both:

- Whether the AR application maintains or improves safety relative to baseline practice
- Whether it provides meaningful improvement in operational performance for the intended task

This dual assessment helps avoid adoption decisions based solely on novelty, speed or feature availability.

5.4 VALIDATE UNDER REPRESENTATIVE CONDITIONS

Augmented reality systems should be evaluated under conditions that are representative

of intended use to the extent practicable. Demonstrations in controlled settings may be useful for early screening, but they are not sufficient to establish suitability for operational deployment where environmental conditions, workload, interruptions, communications constraints and safety demands may be materially different.

Validation should account for relevant operational influences such as:

- Lighting and visibility
- Motion, vibration or weather exposure
- PPE use and physical compatibility
- Noise and communications constraints
- User movement and field-of-view demands
- Expected time pressure and task complexity
- Degraded or abnormal conditions

Where AR is intended for higher-consequence work, validation should extend beyond normal conditions and examine how the system performs when information is delayed, degraded, interrupted or unavailable.

5.5 APPLY A PHASED DEPLOYMENT APPROACH

Implementation should normally proceed in stages rather than as immediate broad adoption. A phased approach allows organizations to confirm expected benefits, identify unanticipated burdens and refine controls before the technology is used in more demanding applications.

A staged approach may include:

1. **Concept or use-case screening**
2. **Limited prototype or simulation-based evaluation**
3. **Pilot use in bounded operational conditions**
4. **Expanded deployment for approved tasks and environments**
5. **Periodic review following operational use**

This progression is especially important where AR changes attention patterns, work sequencing or coordination among personnel. Early deployment should remain bound in scope and should include a clear fallback to the baseline method if the AR system proves unreliable or unsuitable.



5.6 ADDRESS HUMAN FACTORS EXPLICITLY

Human factors should be treated as a primary implementation consideration rather than a secondary usability issue. In operational settings, poor interface design can create direct consequences for workload, hazard perception, communication and task reliability.

Implementation planning should therefore consider:

- Display clutter and information density
- Visual placement and prioritization of content
- Timing and intrusiveness of prompts or alerts
- Ease of user interaction during task execution
- Compatibility with established visual scanning and monitoring behaviors
- Risk of overreliance on prompts or guidance
- Differences in use between experienced and less experienced personnel

Where practical, evaluation should include personnel representative of the actual user population rather than only developers or technical specialists.

5.7 DEFINE OPERATIONAL LIMITATIONS AND REQUIRED CONTROLS

Where AR is found to be suitable only under certain conditions, those limitations should be defined explicitly. Operational limitations may relate to:

- Task type
- User qualification or training level
- Environmental conditions
- Requirement for communications availability
- Exclusion from safety-critical steps
- Use only as a supplemental aid rather than primary information source
- Device maintenance, charging, calibration, or inspection requirements

Required controls should also be specified. Depending on the application, these may include fallback procedures, supervision requirements, alert redesign, simplification of displayed information, communication redundancy or restrictions on use during degraded conditions.

The implementation objective should not be to maximize AR exposure, but to define a scope of use within which the technology has been shown to provide acceptable outcomes.

5.8 DO NOT INFER OPERATIONAL SUITABILITY FROM TRAINING VALUE ALONE

AR may provide significant value for familiarization, procedural rehearsal or competency development. However, training utility should not be treated as equivalent to live operational suitability.

A system that is effective in a lower-consequence training environment may not be suitable for real-time field use where users face time pressure, distractions, environmental variability, moving hazards, communication interruptions or safety-critical decision points.

Where AR is beneficial for training but not validated for field use, that boundary should be stated clearly. Training value is a legitimate implementation outcome, but it is not automatic evidence of field readiness.

5.9 MONITOR PERFORMANCE AND RESIDUAL RISK AFTER DEPLOYMENT

Initial assessment should not be treated as the final basis for continued use. Once deployed, AR applications should be monitored for actual effect on task performance, user behavior, reliability and operational risk.

Post-deployment monitoring may include:

- User feedback on workload and usability
- Observed changes in task duration, error rates or consistency
- Frequency of device or connectivity issues
- Workarounds adopted by users
- Evidence of overreliance or underuse
- Near misses, abnormal events or degraded-condition observations
- Need for software, workflow or alerting changes

This is important because user behavior and task execution may evolve over time in ways not fully visible during initial evaluation. Where operational experience contradicts earlier assumptions, the suitability determination, control measures or approved scope of use should be revisited.



5.10 RECOMMENDATIONS

Based on the considerations presented in this advisory, the following recommendations are offered for the evaluation and implementation of AR in industrial environments in Figure 6.

1

ASSESS AR COMPARATIVELY, NOT ABSTRACTLY
Evaluate the AR-enabled task against the current accepted method of work rather than against generalized expectations of technological benefit.

2

DEFINE THE BASELINE CONDITION EXPLICITLY
Establish the actual current method of task execution before evaluating potential improvement.

3

EVALUATE BOTH SAFETY AND PERFORMANCE
Do not base implementation decisions solely on efficiency, speed or information availability.

6

VALIDATE IN REPRESENTATIVE CONDITIONS
Use testing, pilot deployment or staged evaluation that reflects intended operating conditions and foreseeable degraded scenarios.

5

IDENTIFY TECHNOLOGY-INDUCED HAZARDS AND DEPENDENCIES
Consider distraction, visual competition, latency, registration error, connectivity loss, battery limits and overreliance as part of the operational risk profile.

4

FOCUS ON TASK AND CONTEXT
Suitability depends on the specific task, user population, environment and system configuration.

7

APPLY PHASED DEPLOYMENT WITH BOUNDED SCOPE
Begin with selected tasks and defined limitations rather than broad implementation.

8

TREAT HUMAN FACTORS AS A CORE SAFETY CONSIDERATION
Interface design, workload and situational awareness are central to suitability.

9

DISTINGUISH TRAINING VALUE FROM OPERATIONAL SUITABILITY
Do not infer field-readiness solely from success in learning or rehearsal environments.

10

MONITOR ACTUAL USE AND REVISE AS NEEDED
Continue evaluating performance, user behavior and residual risk after deployment.

Figure 6: Recommendations and considerations for the evaluation and implementation of AR in industrial environments.

5.11 CLOSING OBSERVATIONS

The introduction of AR into industrial work should be viewed as a change to task execution rather than as a simple addition of digital capability. Suitability depends on whether AR improves or at least maintains acceptable safety and operational performance when compared with the existing method of work.

A structured comparative approach helps organizations distinguish between applications where AR provides meaningful value, applications where controls or limitations are required and applications where the technology is not appropriate for the intended use. This provides a more disciplined basis for adoption and supports the responsible integration of AR into industrial operations.



References

ANSI/CAN/UL 8400 *Virtual Reality, Augmented Reality and Mixed Reality Technology Equipment*

ISO/IEC 18023 *Sensor representation in mixed and augmented reality*

ISO/IEC 18039 *Mixed and Augmented Reality*

ISO/IEC 18040 *Live actor and entity representation in mixed and augmented reality (MAR)*

ISO/IEC 18520 *Benchmarking of vision-based spatial registration and tracking methods for mixed and augmented reality (MAR)*

ISO/IEC 23000 *Immersive Stream for XR Overview*

ISO/IEC 5927 *Computer graphics, image processing and environmental data representation — Augmented and virtual reality safety — Guidance on safe immersion, set up and usage*

IEC 62368 *Audio/Video, Information and Communication Technology Equipment*

ISO 9241 *Ergonomics of human-system interaction*

IEEE P3048 Series

Bailey GS, Arruda DG and Stoffregen TA (2022) *Using quantitative data on postural activity to develop methods to predict and prevent cybersickness. Front. Virtual Real. 3:1001080. doi: 10.3389/frvir.2022.1001080*

Bellucci, A., Ruiz, A., Díaz, P., & Aedo, I. (2018, May). *Investigating augmented reality support for novice users in circuit prototyping. In Proceedings of the 2018 international conference on advanced visual interfaces (pp. 1-5).*

Cao, H., Rivera, F.G., Söderlund, H. et al. *Human-centered design of VR interface features to support mental workload and spatial cognition during collaboration tasks in manufacturing. Cogn Tech Work 27, 467–485 (2025). https://doi.org/10.1007/s10111-025-00809-6*

Gerdenitsch, C., Sackl, A., & Hold, P. (2022). *Augmented reality assisted assembly: An action regulation theory perspective on performance and user experience. International Journal of Industrial Ergonomics, 92, 103384.*

Han, X., Chen, Y., Feng, Q., & Luo, H. (2022). *Augmented reality in professional training: A review of the literature from 2001 to 2020. Applied Sciences, 12(3), 1024. https://doi.org/10.3390/app12031024*

Howard, S. (2023). *Visual inspection with augmented reality head-mounted display: An Australian usability case study. Hum. Factors Ergon. Manuf. Serv. Ind., vol. 33, no. 3, pp. 272–296, 2023, doi: https://doi.org/10.1002/hfm.20986.*

Ito, K., Tada, M., Ujike, H., & Hyodo, K. (2021). *Effects of the weight and balance of head-mounted displays on physical load. Applied Sciences, 11(15), 6802. https://doi.org/10.3390/app11156802*

J. H. Lee, K. Yeom and J. -H. Park, “The Effect of Video See-Through HMD on Peripheral Visual Search Performance,” in *IEEE Access*, vol. 11, pp. 85184-85190, 2023, doi: 10.1109/ACCESS.2023.3304363.

Juliano, J. M., Schweighofer, N., & Liew, S. L. (2022). *Increased cognitive load in immersive virtual reality during visuomotor adaptation is associated with decreased long-term retention and context transfer. Journal of NeuroEngineering and Rehabilitation, 19(1), 106.*

Khan, R., Vernooij, J., Salvatori, D., & Hierck, B. P. (2025). *Assessing Cognitive Load Using EEG and Eye-Tracking in 3-D Learning Environments: A Systematic Review. Multimodal Technologies and Interaction, 9(9), 99. https://doi.org/10.3390/mti9090099*

Liberty, J. T., Adamu, P. I., & Bata, S. M. (2026). *From Data to Decisions: How Augmented Reality Transforms Health, Food Safety, and Global Compliance. SN Computer Science, 7(4), 307.*

Marques, B., Ferreira, C., Silva, S. et al. *Exploring different content creation and display methods for remote collaboration supported by extended reality: comparative analysis of distinct task scenarios. Multimed Tools Appl 84, 19785–19815 (2025). https://doi.org/10.1007/s11042-024-19836-y*

Miranda, A., Vallejo, A. M., Ayala, P., Garcia, M. V., & Naranjo, J. E. (2025). *Enhancing industrial processes through augmented reality: A scoping review. Future Internet, 17(8), 358. https://doi.org/10.3390/fi17080358*

Moncur, B., Galvez Trigo, M. J., & Mortara, L. (2023). *Augmented reality to reduce cognitive load in operational decision making. In Proceedings of the International Conference on Human Computer Interaction (LNCS 14019, pp. 328–346). Springer. https://doi.org/10.1007/978-3-031-35017-7_21*

Parasuraman, R., & Riley, V. (1997). *Humans and automation: Use, misuse, disuse, abuse. Human factors, 39(2), 230-253.*

Wang, G., Wang, H.-H., & Huang, Z. (2025). *Enhancing Multi-Region Target Search Efficiency Through Integrated Peripheral Vision and Head-Mounted Display Systems. Information, 16(9), 800. https://doi.org/10.3390/info16090800*



APPENDIX

Selecting Measurement Approaches for Comparative Assessment

Comparative assessment of augmented reality applications may be performed using measurement approaches that vary in complexity, fidelity, resource demand and interpretability. The appropriate approach depends on the purpose of the assessment, the significance of the task, the feasibility of data collection and the level of confidence needed to support the resulting decision.

In many cases, particularly during concept screening, prototype comparison or bounded-use evaluation, relatively simple measures may provide a sufficient basis for comparison. In

other cases, especially where the task is safety-critical or where the mechanisms of performance degradation are not well understood, more instrumented or analytically intensive methods may be appropriate.

For the purposes of this advisory, measurement approaches may be viewed broadly as black-box, gray-box or white-box approaches. These terms are used here as shorthand for the degree to which the assessment relies on directly observable outcomes versus more specialized or higher-fidelity measurement of the mechanisms underlying those outcomes (Figure 7).

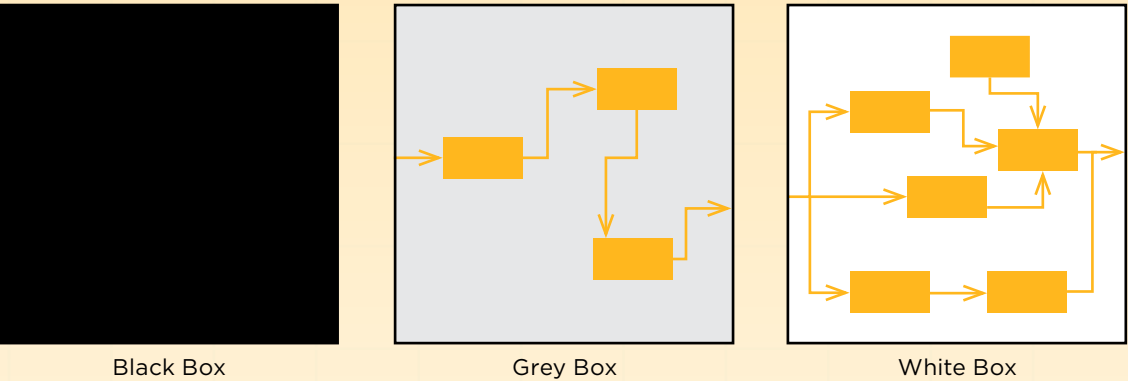


Figure 7: Conceptual comparison of analysis transparency based on data and model interpretability.

A.1 BLACK-BOX APPROACHES

Black-box approaches rely primarily on directly observable or readily captured outcomes. These measures are often the most practical to implement and the easiest to interpret in operational settings.

Examples may include:

- Task completion time
- Task completion rate
- Observed errors or omissions
- Inspection completeness or work quality
- Workload ratings
- Usability feedback
- Near misses, incidents or degraded-function events
- Structured observations of user behavior

Black-box approaches are often well suited to early evaluation, field trials and comparative assessments where practical constraints limit the use of more specialized instrumentation. Their primary advantage is transparency, because the relationship between the measure and the observed task outcome is usually straightforward.

Their primary limitation is that they may reveal that a difference exists without clearly explaining why it exists. For example, an increase in task time or error frequency may be evident, while the underlying cause, such as attentional capture, display clutter or delayed response to environmental cues, may remain uncertain.

A.2 GRAY-BOX APPROACHES

Gray-box approaches provide an intermediate level of insight by combining observable task outcomes with measures that require additional instrumentation, access or analysis, but remain feasible in some controlled or semi-controlled settings.

Examples may include:

- Eye-tracking summary measures
- Biometric indicators
- Device interaction logs
- Response timing data
- Summary measures of head or body movement
- Other derived indicators of attention, workload or user response

Gray-box approaches may help identify changes in workload, attention allocation or interaction behavior that are not fully captured through task outcomes alone. They may therefore provide a useful balance between fidelity and practicality, particularly where the assessment seeks more than surface-level comparison but cannot support highly instrumented analysis.

Their primary limitation is that they often require additional expertise to collect and interpret properly and may still provide only partial visibility into the mechanisms driving performance differences.



A.3 WHITE-BOX APPROACHES

White-box approaches rely on more specialized instrumentation or higher-fidelity analytical methods to examine the mechanisms by which AR affects perception, cognition, motor behavior or task execution.

Examples may include:

- Detailed gaze behavior analysis
- Motion tracking
- Posture or movement analysis
- Reaction-time decomposition
- Physiological workload indicators
- Other instrumented measures intended to provide deeper insight into how users perceive, process and respond during task execution

White-box approaches may be useful where the assessment objective is to identify the mechanism through which AR affect performance or safety, rather than simply determining whether those effects occur. This may be valuable in higher-consequence applications, interface redesign or investigation of observed performance degradations.

Their primary limitations are practical. They may be costly, resource-intensive, difficult to deploy in representative field settings and more demanding to interpret correctly. They may also create a false sense of rigor if the resulting data is not meaningfully connected to the operational decision being made. Advanced instrumentation is not a substitute for judgment. It merely gives judgment more evidence to consider.

A.4 SELECTING AN APPROPRIATE MEASUREMENT APPROACH

The use of more sophisticated measures does not automatically produce a stronger assessment. Measurement approaches should be selected according to the decision the assessment is intended to support and the practical constraints under which it is conducted.

Relevant considerations may include:

- The significance of the task and consequence of error
- Whether the intended use is bounded, operational or safety-critical
- Whether the assessment is being conducted for concept screening, pilot evaluation or deployment decision-making
- Whether the intended task can be evaluated directly or through a representative metatask
- The realism of the evaluation environment
- Access to representative users and required instrumentation
- The interpretability of the resulting data

In many cases, a combination of approaches may be appropriate. Black-box measures may provide a practical baseline for comparison, while gray-box or white-box approaches may be used selectively to investigate specific concerns related to attention, workload, situational awareness or user response.

A.5 LIMITATIONS AND DOCUMENTATION

Regardless of the approach used, limitations in data availability, representativeness, analytical transparency and interpretability should be documented explicitly. This is particularly important where conclusions regarding suitability are based on partial evidence, metatask-based evaluation or highly controlled assessment conditions.

The objective of measurement selection is not to maximize complexity, but to provide evidence that is sufficiently relevant, credible and practical to support a defensible comparative assessment of AR relative to baseline practice.





1701 City Plaza Drive | Spring, TX 77389 USA
1-281-877-6000 | www.eagle.org

© 2026 American Bureau of Shipping.
All rights reserved.