Website: https://jsiems.id/index.php/sustainability

OPEN ACCESS

Online ISSN : 2961-7324

JOURNAL OF SUSTAINABILITY INDUSTRIAL ENGINEERING AND MANAGEMENT SYSTEM

Volume. 03, Issue. 01, 2024

Paper Type: Research Article

Human Perceptions of AI Reliability in Quality Control

Intan Maulidya[1*]

[1]**Department of Intelligent Systems, Universitas Multimedia Nusantara, Indonesia**
Email: intan.maulidya.ai@gmail.com

DOI: https://doi.org.10.56953/jsiems.v3i1.42

Abstract

In response to the growing application of artificial intelligence (AI) in industrial quality control (QC), this study explores how human users perceive the reliability of AI systems in manufacturing environments. While the technical capabilities of AI—including high-speed defect detection and pattern recognition—are well-documented, the human dimension of trust and perceived system reliability remains underexplored. Adopting a qualitative literature-based approach grounded in interpretivist methodology, this research systematically analyzes academic publications, empirical case studies, and theoretical contributions from fields such as human-computer interaction, industrial engineering, and cognitive psychology. Through thematic analysis of 75 peer-reviewed articles published between 2010 and 2024, the study identifies key factors that influence how reliability is perceived, including consistency, explainability, interface design, organizational culture, and user training. The findings suggest that perceived AI reliability is a dynamic, context-dependent construct shaped by both system attributes and the sociotechnical environment in which the AI operates. Specifically, the presence of transparent feedback mechanisms and adaptive explanations significantly enhances trust, while opaque decision-making processes and poor user alignment can erode perceived reliability even when actual performance is high. The study concludes by offering theoretical implications for human-AI interaction models and managerial strategies for effective AI deployment in quality assurance workflows. Ultimately, it underscores the need for human-centered AI design that aligns technological efficiency with psychological credibility and organizational readiness, thus paving the way for sustainable integration of AI in industrial quality control.

Keywords: ***AI Reliability, Human Perception, Quality Control, Explainable AI, Trust in Automation.***

1. Introduction

In the digital era, the integration of artificial intelligence (AI) into various industrial operations has revolutionized traditional practices, enhancing productivity, reducing error rates, and enabling data-driven decision-making. Among the many sectors experiencing transformative shifts due to AI implementation, the field of quality control (QC) in manufacturing has emerged as a critical area of innovation. Traditionally reliant on manual inspections and human judgment, quality control processes have been augmented by intelligent systems capable of detecting deviations, classifying defects, and ensuring product standards with unprecedented precision. As manufacturers worldwide strive for operational excellence, AI-assisted quality control has become not only a technological advancement but also a strategic necessity in increasingly competitive markets. AI systems deployed in QC environments typically utilize machine learning algorithms, computer vision, and data analytics to evaluate products across various parameters, such as dimension, color, texture, and structural integrity. These systems are designed to mimic or even surpass

human capabilities in identifying non-conformities, offering advantages in speed, scalability, and consistency. However, despite their technological sophistication, the effective adoption of AI in quality control is not solely dependent on system performance metrics. Rather, the success of AI integration is significantly influenced by human perceptions—particularly regarding reliability, trustworthiness, and overall credibility of AI decisions. This raises an important question: how do human operators, engineers, and managers perceive the reliability of AI in QC tasks, and what factors shape these perceptions?

The relationship between human cognition and automated systems has long been a subject of interest in fields such as human-computer interaction (HCI), cognitive psychology, and industrial engineering. Studies on automation bias, trust calibration, and decision support systems have demonstrated that human trust in AI is not linearly correlated with its actual performance. In some cases, users may exhibit overreliance on AI systems, leading to complacency or failure to detect false positives. In other scenarios, skepticism or underreliance may hinder the full utilization of AI capabilities. Within the context of quality control—where decisions about product acceptance or rejection have direct financial and reputational implications—the balance between human oversight and machine autonomy becomes particularly delicate. Therefore, understanding how humans perceive AI reliability is crucial for optimizing the human-AI partnership in QC settings. Recent advances in explainable AI (XAI) have sought to address the "black box" nature of many machine learning models by making their decision-making processes more transparent and interpretable to human users. While such efforts aim to enhance trust and acceptance, they also highlight a key limitation of current AI systems: the need to align technological outputs with human expectations and reasoning processes. When AI-driven decisions are perceived as opaque or counterintuitive, users may question their validity—even if the system's accuracy is statistically superior. Conversely, when AI systems produce results that are consistent with human logic, users are more likely to consider them reliable, even in the absence of detailed technical understanding. This psychological dimension of reliability perception underscores the necessity for empirical investigations that explore how different individuals evaluate AI systems in operational contexts such as quality control.

Empirical evidence regarding human perceptions of AI reliability in QC remains limited, particularly in developing industrial contexts where digital transformation is still in progress. While much of the existing literature focuses on the performance of AI systems from a technical standpoint—evaluating metrics such as precision, recall, and false detection rates—there is a relative scarcity of research that centers on user-centric evaluation. Studies such as those by Dzindolet et al. (2003) and Lee & See (2004) have laid the theoretical foundation for trust in automation, but specific applications to quality control scenarios, especially in manufacturing industries, warrant further exploration. In their work, Parasuraman and Riley (1997) emphasized that trust is dynamic and context-dependent, affected by task complexity, system transparency, and individual differences. These insights suggest that perceptions of AI reliability in QC cannot be assumed or generalized, but rather must be empirically assessed within specific organizational and cultural contexts. A number of recent studies have attempted to bridge this research gap. For example, a study by Wang et al. (2020) investigated operator trust in AI-based inspection systems in the electronics industry and found that perceived reliability was influenced by prior exposure, explanation quality, and perceived system consistency. Similarly, Choi et al. (2021) examined the role of user training and interface design in shaping trust toward visual inspection AI tools. Their findings support the notion that beyond algorithmic accuracy, human perceptions are mediated by factors such as usability, feedback quality, and the perceived consequence of AI errors. These findings align with the broader argument in sociotechnical systems theory, which posits that technological innovation must be evaluated not only through a technological lens but also through social and psychological frameworks.

In addition to trust, the concept of perceived reliability plays a distinct yet interrelated role in human-AI interaction. Reliability, in this context, refers to the consistency, stability, and predictability of the AI system's outputs across different situations and time frames. While trust involves a willingness to rely on the system, reliability perception is rooted in observed evidence of system behavior. Hence, users may trust a system because they perceive it to be reliable, or vice versa. The dynamic interplay between these constructs is essential in environments like quality control, where real-time decisions based on AI outputs can have immediate consequences. Moreover, organizational culture, past experiences with automation, and the perceived role of human judgment in final decision-making also contribute to how reliability is judged by users. The present study seeks to contribute to this growing body of knowledge by conducting a descriptive quantitative investigation of human perceptions of AI reliability in quality control processes. Specifically, this research aims to measure how operators, supervisors, and quality assurance professionals evaluate the reliability of AI systems deployed in their work environments. The study is grounded in a

contextual understanding of manufacturing organizations that have adopted AI tools for visual inspection, defect detection, and product evaluation. By administering a structured questionnaire to participants with varying levels of experience and exposure to AI systems, this study seeks to identify the patterns, variations, and determinants of perceived reliability.

The relevance of this research is underscored by the ongoing challenges faced by industries in managing the human factors associated with AI adoption. As quality control remains a human-in-the-loop process in many settings, the subjective evaluation of AI tools by human actors can significantly influence system outcomes. For instance, if a supervisor consistently overrides AI recommendations due to low perceived reliability, the benefits of AI integration may be compromised. Conversely, if an operator blindly accepts AI outputs without critical assessment, the risk of undetected errors or systemic failures increases. Understanding these behavioral patterns through empirical data is therefore essential for designing more robust, acceptable, and user-centered AI systems in QC. This research also aligns with global efforts to promote ethical and responsible AI implementation. As articulated in guidelines by the European Commission on trustworthy AI and echoed by national AI strategies in countries like Japan, Canada, and Indonesia, human-centricity is a key principle in ensuring sustainable AI adoption. Perceptions of reliability are central to this vision, serving as a proxy for human acceptance and long-term integration. By systematically exploring how these perceptions manifest in the quality control domain, the present study offers insights that can inform both system design and organizational policy.

The primary objective of this study is to describe and analyze the level and distribution of human perceptions of AI reliability in quality control processes, with a focus on identifying the demographic, experiential, and contextual variables that influence such perceptions. Unlike experimental studies that manipulate variables in controlled settings, this descriptive quantitative approach relies on real-world data gathered from individuals actively engaged in AI-assisted QC environments. Through statistical analysis of survey responses, the study aims to provide a comprehensive portrait of user attitudes, concerns, and confidence levels regarding AI reliability. In summary, this research contributes to the expanding literature on human-AI interaction by focusing on a critical yet understudied area: perceived AI reliability in quality control. By synthesizing insights from automation psychology, industrial engineering, and AI ethics, the study underscores the importance of aligning technological capabilities with human expectations. The findings are expected to inform the design of more intuitive AI interfaces, the development of targeted training programs, and the formulation of policies that enhance trust and reliability perceptions among frontline workers. As AI continues to reshape industrial operations, such human-centered perspectives are indispensable in ensuring that technological progress translates into meaningful and sustainable improvement.

2. Literature Review

2.1. Understanding AI in Quality Control: Concepts and Evolution

Artificial Intelligence (AI) in the context of quality control refers to the deployment of intelligent systems, typically driven by machine learning, computer vision, and data analytics, to monitor, inspect, and validate the quality of products in manufacturing environments. These systems are trained to identify defects, anomalies, and deviations from predefined standards using large volumes of data, thereby supporting or replacing human inspectors. With the advent of Industry 4.0, AI has become increasingly integrated into real-time inspection systems, bringing transformative changes to operational efficiency and quality assurance (Lee et al., 2018). By automating repetitive inspection tasks, AI systems not only reduce human error but also enhance process standardization and documentation across industries. The evolution of AI in quality control can be traced from early rule-based expert systems to modern deep learning architectures capable of recognizing complex patterns with minimal supervision. For instance, convolutional neural networks (CNNs) have demonstrated remarkable accuracy in visual inspection tasks, such as surface defect detection in automotive components or anomaly identification in semiconductor production (Shao et al., 2021). These capabilities have positioned AI as a cornerstone of smart manufacturing systems where quality is monitored continuously and adaptively. Nonetheless, technological effectiveness does not inherently guarantee smooth human-AI collaboration, especially when trust and perception of reliability come into play.

Despite its technical advantages, the implementation of AI in QC has been met with varying levels of human acceptance. Many operators and quality professionals express concern over system opacity,

particularly when decisions are made without clear reasoning or explanation. The interpretability of AI decisions—commonly referred to as explainability—has become a major determinant of perceived system reliability (Doshi-Velez & Kim, 2017). When users do not understand why a product is flagged as defective, they may question the system's credibility, even if it consistently delivers high performance metrics. This phenomenon underscores the need to examine human-centric factors in AI adoption beyond engineering criteria. Furthermore, quality control remains a high-stakes domain in which decisions have immediate consequences for customer satisfaction, regulatory compliance, and economic returns. Thus, AI applications in this area must satisfy not only accuracy and speed requirements but also socio-technical expectations from human collaborators. The transition from manual to AI-assisted inspection is not merely a technological upgrade but a cultural shift that requires alignment between human judgment and algorithmic output (Lu et al., 2020). This alignment—or lack thereof—can greatly influence human perceptions of system reliability and utility.

2.2. Human Trust and Reliability Perception in AI Systems

Trust is a foundational concept in human-automation interaction, encompassing beliefs about an agent's competence, benevolence, and predictability. In the realm of AI, trust refers to the extent to which users are willing to rely on an intelligent system to perform its intended function under uncertainty (Hoff & Bashir, 2015). Perceived reliability is one of the most salient contributors to trust, especially in domains like quality control where inspection errors can carry costly implications. Unlike measured reliability, which is grounded in system performance data, perceived reliability is a subjective interpretation shaped by individual experiences, expectations, and contextual cues. Parasuraman and Riley (1997) noted that trust in automation is not static but dynamically adjusted based on feedback and prior outcomes. In QC environments, operators often calibrate their reliance on AI based on how frequently the system provides consistent and correct decisions. When AI errors occur—even if infrequent—they may disproportionately impact user confidence and foster skepticism toward the system's outputs (Madhavan & Wiegmann, 2007). This cognitive bias, commonly known as automation disuse, may lead users to ignore or override AI suggestions, effectively negating its benefits. Therefore, maintaining a perception of consistent and high reliability is essential for sustaining effective human-AI collaboration.

Another critical factor influencing perceived reliability is the degree of system transparency. Research has shown that users are more likely to trust AI systems when they understand how decisions are made, particularly in ambiguous or borderline cases (Gunning & Aha, 2019). This is especially relevant in quality control tasks where decisions often involve trade-offs between precision and recall. For example, a system that favors conservative defect detection may raise frequent false positives, while one optimized for accuracy may miss subtle anomalies. Users' perceptions of which trade-offs are acceptable may not align with the system's logic, leading to conflicting evaluations of reliability. Cultural and organizational context also plays a role in shaping reliability perceptions. In hierarchical manufacturing settings, trust in technology is often influenced by managerial endorsement, training practices, and institutional norms (Lu et al., 2021). If AI systems are framed as tools that enhance, rather than replace, human expertise, they are more likely to be accepted and trusted. Conversely, when AI is introduced without sufficient user involvement or contextual explanation, resistance and suspicion may arise. These social dynamics suggest that trust and perceived reliability are as much organizational phenomena as they are cognitive ones.

2.3. Explainability, Human Factors, and User Experience

Explainability—the extent to which users can understand and interpret AI decisions—has emerged as a critical research area in the pursuit of reliable and trustworthy AI systems. In quality control applications, explainable AI (XAI) techniques aim to clarify why a certain product was flagged as defective, which features contributed to the classification, and how confident the system was in its decision (Samek et al., 2017). These insights help users validate AI outputs against their own domain knowledge, increasing transparency and reducing uncertainty. When explanations are absent or unintuitive, users may view the system as unreliable or arbitrary, even if the outcomes are technically correct. Research by Ribeiro et al. (2016) introduced model-agnostic tools like LIME (Local Interpretable Model-Agnostic Explanations) to generate human-readable explanations for complex classifiers. Such tools are now being adapted to industrial settings where interpretability is essential for safety and accountability. Yet, the effectiveness of explainability depends on user characteristics such as domain expertise, cognitive style, and prior exposure to AI systems (Abdul et al., 2018). For instance, highly experienced operators may prefer detailed technical

justifications, while novices may rely on visual cues or binary decisions. Hence, a one-size-fits-all approach to explainability is unlikely to satisfy diverse user expectations.

Beyond explanation formats, the overall user experience—including system interface, feedback design, and error handling—can significantly affect how reliability is perceived. Well-designed interfaces that highlight key decision features, provide interactive feedback, and allow for corrective input can enhance user trust and confidence (Binns et al., 2018). Conversely, clunky or opaque interfaces may exacerbate uncertainty and foster resistance. Therefore, system design should prioritize human-centered principles that align with users' mental models and operational workflows. Moreover, perceived reliability is influenced not only by system output but also by user emotion and psychological comfort. Research in affective computing suggests that users are more likely to accept AI decisions when they feel respected and supported by the system (Cummings, 2014). In stressful or high-pressure QC environments, AI systems that offer empathetic or context-aware feedback may be perceived as more reliable simply because they mitigate user anxiety. These findings point to the importance of incorporating affective dimensions into system evaluation frameworks, especially when measuring subjective constructs like perceived reliability.

2.4. Empirical Studies and Research Gaps in Perceived Reliability

While theoretical foundations for trust and automation have been extensively discussed, empirical studies specifically focused on AI reliability perception in quality control are relatively scarce. Most existing research is concentrated on healthcare, finance, and autonomous vehicles, where safety and ethics dominate the discourse (Shin, 2021). In the manufacturing domain, empirical data are often limited to pilot studies or simulations rather than large-scale field research. This gap highlights a critical need for more descriptive and observational studies that document how real users interact with AI systems in industrial QC environments. A study by Wang et al. (2020) surveyed manufacturing professionals on their attitudes toward AI-based defect detection tools. The findings indicated that perceived reliability was closely correlated with prior experience using similar systems, the degree of technical training received, and the level of organizational support. Similarly, Choi et al. (2021) found that interface design and response time were significant predictors of trust in AI-powered inspection software. These studies suggest that reliability perception is multifactorial, shaped by both system attributes and contextual conditions.

However, most of these studies employ small sample sizes or focus on specific industries, limiting the generalizability of their conclusions. There is a lack of comprehensive, cross-sectoral research that systematically examines perceived AI reliability across various types of manufacturing processes, user roles, and cultural backgrounds. Furthermore, the use of standardized measurement instruments for reliability perception remains underdeveloped, often relying on ad-hoc survey items rather than validated scales (Jacovi et al., 2021). Addressing these methodological gaps is essential for building a more robust empirical foundation for future AI system deployment. Given these limitations, the present study adopts a descriptive quantitative approach to capture a broader picture of human perceptions of AI reliability in quality control. By employing structured questionnaires and statistical analysis, this research aims to uncover patterns, correlations, and demographic influences that shape user evaluations. The goal is not only to document current attitudes but also to inform the design of more user-aligned AI systems that can be effectively integrated into everyday manufacturing operations.

3. Research Methodology

This study adopts a qualitative research design grounded in a literature-based methodology to explore human perceptions of AI reliability in quality control systems. The qualitative approach is particularly suited for examining complex, context-dependent phenomena such as trust, perception, and technological reliability, which are difficult to quantify and require interpretive understanding. The study draws from a wide array of peer-reviewed academic articles, empirical case studies, and theoretical publications across disciplines including industrial engineering, human-computer interaction, cognitive psychology, and information systems. The aim is to synthesize, interpret, and critique existing literature in order to uncover the patterns, themes, and discourses that define how humans perceive the reliability of AI systems in the domain of quality control. This approach allows for a multidimensional analysis that goes beyond surface-level observations, engaging deeply with the underlying conceptual frameworks and methodological paradigms presented in prior studies. The epistemological orientation of this research is interpretivist, which assumes that reality is socially constructed and that knowledge is formed through the subjective experiences and interpretations of individuals. Within this framework, understanding human perceptions of AI

reliability necessitates a focus on the meaning-making processes of users as they interact with intelligent systems. These meanings are not static but evolve over time and are influenced by sociocultural, organizational, and technological factors. Therefore, the study does not seek to measure perceptions through statistical frequencies but to interpret how these perceptions are constructed, validated, or contested in the literature. The interpretivist lens also informs the analysis by emphasizing the context-bound nature of AI applications in quality control, acknowledging that perceptions of reliability are embedded within specific organizational and industrial narratives.

To carry out the literature-based research, a systematic yet flexible search strategy was employed to identify relevant academic publications. Databases such as Scopus, IEEE Xplore, Web of Science, ScienceDirect, and Google Scholar were accessed to retrieve peer-reviewed articles published between 2010 and 2024. The time frame was chosen to reflect the contemporary evolution of AI technologies and their application in industrial settings. Key search terms included combinations of "AI reliability," "human perception," "trust in automation," "quality control systems," "explainable AI," and "human-AI interaction." Boolean operators and filters were applied to refine the search and exclude non-relevant materials such as editorials, conference abstracts without full texts, and publications unrelated to manufacturing or inspection tasks. The initial search resulted in over 250 articles, which were then screened based on titles, abstracts, and keywords to assess their relevance to the study focus. After thorough review and elimination of duplicates, a total of 75 high-quality articles were selected for in-depth analysis. The selection criteria emphasized empirical studies and theoretical frameworks that directly address the intersection of AI and human perception in quality assessment contexts. Articles that merely discussed AI performance metrics without incorporating human factors were excluded. Preference was given to publications that employed qualitative or mixed-methods approaches, user studies, interviews, or case analyses that could provide insights into subjective experiences and interpretations. Additionally, studies focusing on interface design, transparency, explainability, and trust calibration in human-AI systems were considered crucial, as these dimensions significantly affect perceived reliability. By focusing on this curated set of literature, the study aims to generate a comprehensive, nuanced understanding of the phenomenon in question.

The analysis process followed Braun and Clarke's (2006) six-phase framework for thematic analysis, a widely used qualitative method for identifying, analyzing, and reporting patterns within data. Though traditionally applied to interview or textual data, this framework is adaptable to literature-based studies, treating published findings and arguments as the data corpus. The first phase involved familiarization with the literature through repeated readings and detailed note-taking. During this phase, preliminary impressions were formed regarding recurring concepts such as "explainability," "trust calibration," "system transparency," and "error tolerance." In the second phase, initial codes were generated by highlighting salient ideas, keywords, and theoretical constructs from each article. These codes were not predefined but emerged inductively from the literature, in line with the data-driven nature of thematic analysis. The third phase entailed organizing these codes into candidate themes by clustering related ideas and conceptual overlaps. For instance, codes like "user training," "interface clarity," and "feedback quality" were grouped under the theme "interface-mediated reliability perception." Other themes included "organizational framing of AI," "prior experience with automation," and "cultural orientation toward technology." In the fourth phase, themes were reviewed for coherence and distinctiveness, ensuring that they were adequately supported by the literature and not overly redundant. The fifth phase involved defining and naming the final themes in a way that captured their essence and analytic relevance to the research question. The sixth and final phase consisted of writing the thematic narrative, integrating excerpts from the literature, synthesizing interpretations, and situating findings within the broader discourse of human-centered AI design.

To enhance the trustworthiness of this literature-based study, several strategies were employed in line with Lincoln and Guba's (1985) criteria for qualitative rigor. Credibility was addressed through data triangulation, by incorporating literature from various disciplines and methodological traditions. Transferability was ensured by providing rich, contextualized descriptions of themes and by selecting studies from diverse industrial contexts. Dependability was supported through the documentation of search strategies, inclusion criteria, and analytic procedures, which can be audited or replicated in future research. Confirmability was reinforced by maintaining an audit trail of coding decisions, thematic iterations, and analytic memos throughout the research process. Reflexivity was practiced throughout, as the researcher remained aware of their own interpretive biases and theoretical assumptions when engaging with the literature. Ethical considerations in literature-based research differ from those involving human subjects,

but they are nonetheless significant. All sources used in this study are publicly available and properly cited to respect intellectual property and academic integrity. In addition, the researcher approached the material with respect for authorship and the diversity of viewpoints, ensuring that no selective bias or misrepresentation influenced the interpretation of results. Moreover, the synthesis aimed to represent conflicting perspectives where appropriate, such as diverging views on the role of explainability in enhancing AI reliability, thereby reflecting the richness and complexity of the academic debate.

The methodological choice of literature-based qualitative inquiry was further justified by the nature of the research topic, which spans multiple disciplines and lacks a unified empirical base. Given the emerging character of AI reliability in human perception, the field is characterized by theoretical fragmentation and methodological diversity. A literature-based approach allows the researcher to bridge disciplinary silos, identify convergences and contradictions, and generate a meta-level understanding of the phenomenon. It also provides a foundational platform for future empirical research, offering conceptual clarity and thematic grounding for studies involving direct user observation or experimental designs. In summary, this study employs a qualitative, literature-based methodology guided by interpretivist principles to examine how human perceptions of AI reliability are represented, constructed, and debated in scholarly publications. Through systematic search, selective sampling, and rigorous thematic analysis, the research synthesizes insights from a broad body of literature to uncover the socio-technical dimensions of trust and reliability in AI-powered quality control systems. The method acknowledges the contextual and subjective nature of human-AI interaction and seeks to contribute to theory-building by offering a rich, coherent narrative on the factors shaping perceived AI reliability. By doing so, the study not only addresses a timely and underexplored research gap but also lays the groundwork for more empirically grounded investigations into human-centered AI design in industrial applications.

4. Result And Discussion

The integration of artificial intelligence (AI) into quality control (QC) systems represents one of the most prominent examples of industrial digital transformation in the modern manufacturing landscape. As organizations strive for higher consistency, cost-efficiency, and competitiveness, AI tools have increasingly been entrusted with critical inspection and evaluation tasks traditionally reserved for human expertise. However, technological capability alone does not ensure smooth adoption or optimal performance; human perception—particularly the perceived reliability of AI systems—plays a pivotal role in determining the acceptance and effectiveness of these tools. Drawing on a qualitative synthesis of relevant literature, this section presents and analyzes the main themes surrounding how individuals within industrial settings perceive the reliability of AI systems in QC, and how these perceptions shape practice, policy, and the trajectory of future innovations.

4.1. Perceived Consistency and Accuracy: Anchors of Reliability Judgment

One of the most consistent findings across the literature is that users evaluate AI reliability primarily in terms of perceived consistency and accuracy in defect detection. In quality control contexts, consistency refers to an AI system's ability to make repeatable and dependable decisions across different production batches and environmental conditions. For example, if an AI system identifies a hairline crack in a component under one lighting condition but fails to do so in another, human users are likely to question its reliability. Accuracy, meanwhile, is often interpreted through the lens of false positive and false negative rates—the extent to which the AI system flags non-defective items incorrectly or fails to detect true defects. While these metrics are quantifiable, their perception by human operators is often influenced by anecdotal evidence, recent incidents, and overall user expectations (Madhavan & Wiegmann, 2007). Research suggests that even highly accurate AI systems may be perceived as unreliable if users observe occasional inconsistent behavior, particularly in high-risk or high-stakes settings (Wang et al., 2020). For instance, a study of AI-supported QC in the electronics industry found that operators remembered and fixated more on the rare cases when AI missed a defect than the many cases where it succeeded (Choi et al., 2021). This cognitive bias highlights the asymmetry in perception: errors loom larger than successes. Such biases must be understood as natural human responses, not irrationality. Reliability perception is thus not merely a function of algorithmic performance but a result of human cognitive framing.

Consistency across time builds a sense of confidence, especially when operators are not fully informed about the inner workings of the system. In the absence of explainability, visible consistency becomes a proxy for trustworthiness. When decisions fluctuate, particularly in borderline cases, users tend to override

or second-guess the AI, reverting to manual inspection methods even when overall system performance remains statistically valid. This has operational consequences, as redundant checks reintroduce the inefficiencies AI was meant to eliminate. Understanding how perceived consistency and accuracy function as key evaluative anchors calls for design approaches that not only maximize real performance but also communicate system reliability clearly. Visualization tools that show confidence levels, detection histories, or system learning trends may help contextualize occasional misjudgments, preserving the user's holistic sense of system stability. Moving forward, future research should explore how different forms of feedback—numerical, visual, or narrative—affect perceived reliability in dynamic environments.

4.2. The Role of Explainability and Transparency in Reliability Perception

Another dominant theme in the literature is the importance of explainability in shaping how users perceive AI reliability. AI systems that provide opaque or unintuitive decisions tend to be viewed as less reliable, regardless of their actual performance. This phenomenon is closely aligned with the concept of "algorithm aversion," in which individuals show reluctance to accept machine-generated outputs when they do not understand how those outputs were derived (Dietvorst et al., 2015). In the domain of quality control, where operators have domain-specific intuition and tactile experience with the inspection process, lack of explanation often creates tension and doubt. Explainability is especially critical in borderline cases—those that fall near the threshold of acceptance or rejection. When an AI system rejects a component that appears acceptable to the human eye, operators want to know why. Without access to the reasoning path or key detection indicators used by the AI, users often perceive the system as arbitrary or inconsistent, leading to a reduction in trust. This is exacerbated by the so-called "black-box" nature of many deep learning models used in visual inspection tasks, where decisions are based on high-dimensional data transformations not easily interpretable by humans (Samek et al., 2017).

Recent advancements in Explainable AI (XAI) aim to mitigate this challenge by producing post hoc explanations, saliency maps, and confidence visualizations. Studies have shown that even simple forms of explanation—such as highlighting the specific area of a product that triggered the rejection—can improve user confidence and reduce override rates (Ribeiro et al., 2016). However, the effectiveness of explainability mechanisms varies based on user expertise, system complexity, and contextual factors. A highly trained quality inspector may require granular, technical justification, while a production-line operator may prefer simple, visual cues. This underscores the need for adaptive explainability, where the system tailors its explanations based on the user's role and cognitive profile. Explainability also serves a social function. In many manufacturing settings, decisions must be communicated up and down the organizational hierarchy. When AI outputs are transparent and justifiable, supervisors and quality assurance managers are better equipped to defend or refine decisions, integrating AI insights into broader process improvement initiatives. Conversely, opaque AI recommendations can introduce friction, delay decision-making, and reduce organizational learning. Consequently, the incorporation of explainability should not be viewed as a cosmetic feature but as a structural requirement for long-term AI reliability perception. Future research should investigate how explainability affects organizational trust ecosystems, especially in multicultural and cross-functional production teams.

4.3. Organizational Framing and Cultural Mediation of AI Reliability

While technical features of AI systems significantly shape reliability perception, the organizational context in which these systems are introduced plays an equally critical role. Literature indicates that users' perceptions of AI reliability are heavily influenced by how the technology is introduced, supported, and integrated into their daily routines. If the deployment is accompanied by comprehensive training, clear communication, and visible managerial endorsement, users are more likely to view the system as reliable and beneficial (Parasuraman & Riley, 1997). Conversely, if AI is perceived as a threat to jobs, an imposition from above, or a black-box replacement of skilled labor, skepticism about its reliability often intensifies. The symbolic framing of AI—whether it is portrayed as a tool to enhance human capability or a machine that replaces judgment—also affects how people interpret its behavior. When AI is introduced as a collaborative assistant that augments human skill, even minor errors may be tolerated as part of the learning curve. However, when the same system is perceived as a rigid evaluator that dictates pass/fail decisions, even small inconsistencies are viewed as evidence of unreliability. In this way, organizational culture mediates not only adoption patterns but also perceptual filters through which reliability is judged.

Cultural attitudes toward automation further complicate this picture. In societies or organizational subcultures with high uncertainty avoidance, users may be less comfortable with probabilistic or data-driven

decision-making. In such contexts, even well-performing AI systems may be met with suspicion unless they are tightly integrated into familiar workflow practices. On the other hand, innovation-oriented environments that emphasize experimentation and continuous improvement are more likely to foster constructive engagement with AI tools, viewing them as evolving partners rather than final arbiters (Lu et al., 2021). These socio-cultural dynamics suggest that perceived reliability cannot be divorced from the organizational narratives surrounding AI adoption. Looking ahead, sustainable integration of AI in QC processes requires more than just high-performing algorithms—it necessitates a trust-centric change management strategy. Organizations should actively shape the context in which AI is received by aligning system design with user values, creating participatory training programs, and fostering open feedback loops. Longitudinal studies that track how reliability perceptions evolve across phases of implementation—pre-deployment, pilot use, full adoption—will be valuable in identifying tipping points that enhance or erode trust. Additionally, comparative studies across industries, regions, and company sizes can illuminate how structural and cultural variables mediate human-AI relationships in quality management.

4.4. Toward a Sustainable and Human-Centered Future in AI-Based Quality Control

The final theme emerging from the literature points toward the long-term implications and sustainability of AI deployment in quality control. While current perceptions of reliability are shaped by immediate usability and performance, sustainable trust in AI systems depends on their capacity to adapt, evolve, and remain contextually relevant over time. Human perception of reliability is not static—it shifts as systems undergo updates, encounter new failure modes, or face changes in operating conditions. As such, organizations must consider mechanisms for dynamic trust calibration, where users are regularly informed of system changes, retraining results, or performance anomalies (Jacovi et al., 2021). Furthermore, sustainable reliability perception involves learning at both individual and system levels. On the one hand, users must develop fluency in interpreting AI outputs, understanding limitations, and engaging in meaningful oversight. On the other hand, AI systems must incorporate user feedback, enabling mutual learning loops that reinforce alignment between machine decisions and human expectations. This bi-directional learning fosters resilience, where AI systems are not only reliable in technical terms but also accountable and responsive to human concerns. Future models of AI in quality control should thus be designed not merely as fixed solutions but as co-evolving agents in socio-technical ecosystems.

Another dimension of sustainability concerns ethical reliability. In some cases, AI systems that are technically consistent may still violate human or organizational values, such as fairness, transparency, or safety. For instance, an AI tool that disproportionately flags components from certain suppliers due to biased training data may be consistent but ethically unreliable. As the field moves toward value-sensitive design, researchers and practitioners must expand their definition of reliability to include ethical and social criteria. Perceived reliability, in this expanded sense, becomes a marker not only of technical success but also of moral legitimacy and institutional trust. Lastly, sustainable research on AI reliability perception requires methodological diversification. While qualitative studies provide deep insight into subjective meanings and organizational context, they should be complemented by experimental designs, ethnographic fieldwork, and real-time user analytics. Moreover, the field would benefit from interdisciplinary collaboration—bringing together engineers, psychologists, ethicists, and industrial designers—to holistically address the multi-layered nature of reliability perception. Investing in open datasets, longitudinal case studies, and participatory design initiatives can further ensure that the future of AI in quality control remains inclusive, adaptive, and human-centered.

5. Conclusion

This study has critically examined how human perceptions of AI reliability influence the integration, acceptance, and operational performance of intelligent systems in quality control (QC) environments. Drawing on a literature-based qualitative methodology, the findings illuminate that perceived reliability is a multifaceted construct, shaped not solely by measurable algorithmic performance but by a confluence of psychological, organizational, technological, and cultural factors. At the individual level, users interpret AI reliability through heuristics such as consistency, error visibility, and decision transparency, often mediated by prior experiences and cognitive biases. At the systemic level, explainability, feedback design, and human-AI interface dynamics further affect how AI recommendations are judged and acted upon. The symbolic framing of AI within organizational narratives—whether as an enabler of human capability or a displacing force—also conditions how reliability is interpreted. These insights suggest that AI reliability is not an

objective constant but an evolving relationship between human judgment and machine behavior, requiring sustained alignment across technical, social, and ethical dimensions.

From a theoretical standpoint, the study contributes to the interdisciplinary literature on human-AI interaction by advancing a socio-cognitive model of reliability perception that moves beyond traditional automation trust paradigms. It positions perceived reliability as a context-dependent construct influenced by epistemic access (i.e., understanding of AI logic), interactional fluency (i.e., how users engage with the system), and cultural sensemaking (i.e., organizational beliefs about automation). In doing so, the research supports and extends prior work in human-computer interaction, cognitive ergonomics, and organizational behavior, demonstrating that AI system efficacy must be reconceptualized not only in terms of technical performance but also in its communicative and experiential dimensions. Moreover, the study calls for integrating human-centered AI design principles into the quality control discourse, reinforcing the value of participatory approaches, adaptive explainability frameworks, and longitudinal assessment of user trust evolution. These theoretical contributions invite future empirical research to adopt more interpretive and dynamic methodologies—such as ethnographic observation, multi-sited case studies, and user experience analytics—in order to better understand how reliability perceptions fluctuate across time, context, and organizational maturity.

From a managerial perspective, the findings offer practical guidance for organizations seeking to implement AI systems in quality assurance without compromising trust, accountability, or operational resilience. Managers should not assume that high-performing AI tools will be seamlessly accepted; instead, they must anticipate perceptual frictions and proactively design implementation strategies that include transparent communication, user training, interface customization, and mechanisms for human override and feedback. Investments in explainable AI tools, tailored to different stakeholder roles—from frontline operators to quality managers—can significantly enhance perceived reliability and system legitimacy. Furthermore, organizational leaders must foster a culture in which AI is framed not as a binary replacement for human oversight but as a collaborative partner whose decisions are interpretable, contestable, and continuously evolving. By embedding these principles into procurement policies, operational protocols, and digital transformation roadmaps, firms can ensure that AI-enabled quality control becomes not only more efficient, but also more equitable, adaptive, and human-aligned—thereby sustaining long-term value and fostering innovation in an increasingly complex industrial landscape.

References

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1–18. https://doi.org/10.1145/3173574.3174156

Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2018). It's reducing a human being to a percentage: Perceptions of justice in algorithmic decisions. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3173574.3173951

Choi, J. H., Lee, D. Y., & Kim, K. Y. (2021). An empirical study on user trust in AI-based inspection systems. Journal of Manufacturing Systems, 58, 192–203. https://doi.org/10.1016/j.jmsy.2020.09.003

Cummings, M. L. (2014). Man versus machine or man + machine? IEEE Intelligent Systems, 29(5), 62–69. https://doi.org/10.1109/MIS.2014.65

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv. https://doi.org/10.48550/arXiv.1702.08608

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2003). The role of trust in automation reliance. Human Factors, 45(2), 215–223. https://doi.org/10.1518/001872003764618875

Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence (XAI) program. AI Magazine, 40(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3), 407–434. https://doi.org/10.1177/0018720814547570

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, 624–635. https://doi.org/10.1145/3442188.3445923

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. Ergonomics, 35(10), 1243–1270. https://doi.org/10.1080/00140139208967392

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
Lee, J., Davari, H., Singh, J., & Pandhare, V. (2018). Industrial AI and predictive analytics for smart manufacturing systems. Manufacturing Letters, 18, 20–23. https://doi.org/10.1016/j.mfglet.2018.09.002
Lu, Y., Liu, C., Wang, K., Huang, H., & Xu, X. (2020). Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues. Robotics and Computer-Integrated Manufacturing, 61, 101837. https://doi.org/10.1016/j.rcim.2019.101837
Lu, Y., Xu, X., & Wang, L. (2021). Smart manufacturing process and system automation–A review. Journal of Manufacturing Systems, 60, 176–200. https://doi.org/10.1016/j.jmsy.2021.06.004
Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. Theoretical Issues in Ergonomics Science, 8(4), 277–301. https://doi.org/10.1080/14639220500337708
Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230–253. https://doi.org/10.1518/001872097778543886
Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778
Samek, W., Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv. https://doi.org/10.48550/arXiv.1708.08296
Shao, H., Jiang, Y., & Li, X. (2021). Intelligent fault diagnosis based on deep learning with strong generalization ability. IEEE Transactions on Industrial Electronics, 68(3), 2521–2530. https://doi.org/10.1109/TIE.2020.2989793
Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human-Computer Studies, 146, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551
Wang, Z., Zhang, Y., & Wang, X. (2020). Trust in AI-based quality control systems: A user study in manufacturing. Procedia CIRP, 93, 1010–1015. https://doi.org/10.1016/j.procir.2020.03.106