

Article

Resilient Information Quality in Social Media Environments: A Framework for Evaluating Information Under Misinformation and Algorithmic Amplification

Ana Raquel Azevedo ^{1,2,*} , Fátima Gonçalves ¹  and Joaquim Brigas ¹¹ IMEDIALAB, Polytechnic of Guarda, 6300-559 Guarda, Portugal; fgoncalves@ipg.pt (F.G.); joaquimbrigas@ipg.pt (J.B.)² DEGEIT, University of Aveiro, 3810-193 Aveiro, Portugal

* Correspondence: raquel.azevedo@ipg.pt

Abstract

Contemporary information environments have transformed how information is produced, circulated, and evaluated, especially in social media and platform-mediated communication contexts. Although Information Quality (IQ) and misinformation research has provided important criteria for assessing accuracy, completeness, relevance, credibility, and timeliness, existing approaches offer limited insight into how information preserves meaning, credibility, and contextual integrity under decontextualization, emotional amplification, algorithmic visibility, and strategic manipulation. Addressing this gap, this study introduces Resilient Information Quality (RIQ) as a conceptual and analytical framework for evaluating information resilience in digital communication. An exploratory framework-development design is adopted. The framework integrates six dimensions: Manipulability Resilience, Contextual Transparency, Interpretability, Emotional Framing, Audience Resilience, and Platform Suitability. These dimensions are operationalized through a structured scoring rubric and applied to fourteen analytically constructed communication samples derived from recurring real-world information patterns. The findings classify samples into critical, vulnerable, and robust configurations. Emotionally charged and low-transparency messages show high engagement potential but weak resistance to distortion, while institutional and evidence-based communication shows stronger resilience through contextual grounding, source transparency, and structural coherence. The study concludes that IQ should be assessed not only through correctness, but also through manipulation resistance, contextual credibility, and safe platform adaptation.

Keywords: information quality; resilient information quality; misinformation; disinformation; media literacy; digital communication



Academic Editor: Inês Amaral

Received: 20 May 2026

Revised: 3 July 2026

Accepted: 6 July 2026

Published: 8 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Contemporary digital environments are increasingly shaped by algorithmic visibility, rapid information flows, emotional engagement, and fragmented patterns of content consumption, particularly within social media platforms. Studies have shown that even highly connected users often struggle to critically evaluate information, identify reliable sources, or recognize manipulative content, highlighting the growing importance of contextual interpretation, critical evaluation, and platform dynamics in determining how information is perceived and trusted (Rodríguez Castillo et al., 2026).

IQ has long been regarded as a central construct in information systems, knowledge management, and communication research. It is widely understood as a context-dependent and process-oriented concept, shaped by user needs, organizational processes, technological systems, and the environments in which information is produced and used (Eppler, 2006). Traditional IQ models have commonly emphasized dimensions such as accuracy, completeness, timeliness, relevance, consistency, and accessibility. These dimensions were originally developed for relatively stable information settings, where data were produced, stored, and evaluated within controlled organizational or technical environments and where reliable reference points could be clearly established (Wang & Strong, 1996; Eppler & Wittig, 2000; Lee et al., 2002).

Building on these foundations, subsequent studies have advanced multidimensional approaches to IQ assessment. Frameworks such as the AIM Quality Methodology (AIMQ) (Lee et al., 2002), together with related work on understandability, usefulness, and data management, have shown that IQ cannot be reduced to a single attribute but must instead be assessed according to the specific purpose, users, and context of use (Eppler & Wittig, 2000; Lee et al., 2002; Strong et al., 1997). Alongside developments in IQ research, media literacy initiatives have emphasized strengthening users' critical evaluation competencies and the ability to recognize misinformation (Herrero-Diz & López-Rufino, 2021), whereas fact-checking approaches have focused on the verification of factual claims and the mitigation of misinformation (Nyhan et al., 2019).

In this sense, IQ has progressively moved from a purely technical concern toward a broader evaluative construct concerned with how information supports interpretation, decision-making, and communication.

However, the assumptions on which many traditional models of IQ were built are being challenged by the modern digital information environments. Although access to online information has expanded dramatically, evaluating the quality and reliability of digital content remains challenging for many users, reinforcing the need for more robust approaches to information assessment (Battineni et al., 2020).

Information is not merely stored or transmitted in platform-mediated spaces but is constantly remade through algorithmic visibility, social sharing, emotional framing, visual presentation, and user interpretation. This challenge has become especially visible with the growth of misinformation and disinformation, where misleading or manipulative content can spread rapidly without any evidence to support it (Wardle & Derakhshan, 2017; Lewandowsky et al., 2017). IQ here is not just a matter of the factual accuracy of the information, but also of how it is framed, perceived, amplified and possibly distorted.

This issue is particularly crucial because even accurate information can be disbelieved, misinterpreted, dismissed, or manipulated in emotionally charged and polarized communication contexts. Susceptibility to false information has been associated with limited analytical thinking, and engagement and sharing can be encouraged by emotional content, cognitive ease, simplicity, and visual appeal (Pennycook & Rand, 2019; Vosoughi et al., 2018). In this way, misinformation can spread faster and wider, not just because it is false but also because it is often easier to process, emotionally appealing, and strategically calibrated to the dynamics of digital platforms (Vosoughi et al., 2018). These conditions imply that IQ cannot be measured by intrinsic qualities alone such as correctness or completeness.

Research in communication further supports this broader understanding of IQ. Classical persuasion theory has demonstrated that the effectiveness of communication is a function of the audience, the message and the source (Hovland et al., 1953). Likewise, research on the organizational and social use of information has shown that the value

and effectiveness of information for organizations depend on how it is interpreted and integrated into specific contexts (Zárraga-Rodríguez & Álvarez, 2014). Research on media literacy, contextual evaluation, and fact-checking shows that the ability to resist misinformation depends on users' ability to evaluate credibility, recognize manipulation, interpret information in context, and verify factual claims (Herrero-Diz & López-Rufino, 2021; Wardle & Derakhshan, 2017). These contributions suggest that IQ must take into account both the structural properties of information and the conditions in which information is received, transformed, and acted upon.

Despite these advances, important limitations remain. First, many traditional IQ models still tend to favour relatively static properties of information, such as accuracy, completeness, and timeliness, but pay less attention to how information behaves under conditions of decontextualization, emotional amplification, and strategic manipulation. Second, research on IQ, misinformation, media literacy and communication has often developed in parallel, leading to limited integration of information systems perspectives with communication-oriented approaches. Third, existing frameworks do not provide adequate conceptual tools to assess the capacity of information to retain its clarity, credibility, and meaning under the conditions of platform-mediated diffusion, algorithmic amplification, and adversarial reinterpretation.

This gap is particularly significant in contemporary public and institutional communication. Scientific, organizational and institutional information may be accurate but may still fail to generate trust or understanding when it is perceived as too technical, inaccessible, elitist or disconnected from public concerns. Therefore, communication scholars have underscored the importance of clarity, credibility signalling, comprehensibility, emotional responsiveness and audience-oriented communication to enhance public engagement and resilience (Scheufele, 2014; Kahan, 2017; Jamieson, 2018). These contributions highlight the need for a more nuanced definition of high-quality information, considering not only the factual reliability of information but also its ability to withstand fragmentation, polarization and manipulation in today's media environment (Wardle, 2018).

Addressing this gap, the present study introduces RIQ as a conceptual and analytical extension of traditional IQ frameworks. RIQ is proposed to capture the capacity of information to maintain its clarity, credibility, contextual integrity, and intended meaning under conditions of manipulation, emotional amplification, and platform-mediated dissemination. Rather than replacing established IQ dimensions, the proposed framework integrates and extends existing concepts from IQ research, source credibility, media literacy, framing theory, platform studies, and misinformation research into a resilience-oriented model. Its contribution lies not in treating these concepts as isolated criteria, but in organizing them into a unified framework for assessing how information resists distortion across message, audience, and platform conditions.

Specifically, this study is guided by three research questions: (RQ1) Which dimensions are required to evaluate information resilience beyond traditional IQ attributes? (RQ2) How can these dimensions be operationalized into a structured analytical framework? (RQ3) How does the RIQ framework distinguish between critical, vulnerable, and robust communication patterns? In addressing these questions, the study develops RIQ as a conceptual and analytical framework, operationalizes it through six dimensions, and applies it to diverse communication samples in an exploratory manner. By doing so, it contributes to IQ research by integrating insights from information systems, communication theory, misinformation studies, platform studies, and media literacy into a unified framework for evaluating how information maintains credibility and meaning in platform-mediated environments.

The remainder of this article is organized as follows. Section 2 reviews the literature on misinformation, emotional and communicative dynamics, platform logic, AI-generated misinformation, and the need to rethink traditional IQ models. Section 3 presents the materials and methods, including the research design, sample selection, conceptual structure of the RIQ framework, operationalization of its six dimensions, scoring procedure, inter-rater reliability assessment, and classification framework. Section 4 reports the results of the case-based, comparative, and integrated analyses. Section 5 discusses the main findings and theoretical and practical implications. Finally, Section 6 summarizes the main contributions of the study, acknowledges its limitations, outlines future research directions, and concludes the article.

2. Literature Review

Misinformation refers to false or misleading information shared without intent to deceive, often by individuals who believe it to be true. It differs from disinformation, which involves intentional deception, and from malinformation, where factual content is used out of context to cause harm (Wardle & Derakhshan, 2017). While this study focuses primarily on misinformation, the analysis acknowledges overlaps with disinformation, particularly in cases where intent is ambiguous, such as AI-generated content. In practical terms, misinformation and disinformation share common communicative aspects (emotionally framed, simplified narratives, platform-driven dissemination) that inform the framework described in this study.

The rapid transmission of false information through digital networks, driven by emotional, simplistic, and algorithmically amplified characteristics, is one of the most significant current threats to information integrity. However, misinformation reflects not only false content, but a broader breakdown in how information is produced, transmitted, and interpreted. IQ models often do not fully account for how information functions in real-world situations, and especially highly emotional, partisan contexts driven by platforms (Wardle & Derakhshan, 2017; Lewandowsky et al., 2017).

This section focuses on both the structural and psychological factors that enable misinformation not just to emerge, but also to be persuasive and often preferable. This involves the use of emotions in the structure of the message itself (fear, outrage, or the sense of belonging); the aesthetic qualities used in the production of misinformation (humanization of messages), as well as the structure of platforms, which favours speed, virality, and engagement over clarity and deliberation.

Accordingly, the literature review is organized around four complementary perspectives that support the development of the RIQ framework. First, emotional and communicative dynamics are examined because misinformation often gains influence through affective resonance, accessible formats, and persuasive framing rather than factual strength alone. Second, algorithmic and platform logics are reviewed because the visibility and circulation of information are increasingly shaped by engagement-oriented infrastructures that may amplify fragile or misleading content. Third, the role of AI is considered because generative technologies intensify both the production of synthetic misinformation and the possibilities for detection, verification, and media literacy support. Finally, these strands are brought together to explain why traditional IQ models must be reconsidered in light of manipulation, decontextualization, audience vulnerability, and platform-mediated diffusion. This structure therefore provides the conceptual basis for moving from conventional IQ toward a resilience-oriented framework.

2.1. Emotional and Communicative Dynamics of Misinformation

The spread of misinformation is often driven more by emotional engagement than by factual accuracy. Emotionally charged messages that evoke fear, anger, or surprise tend to be more memorable, more easily recalled, and more likely to be shared, regardless of their veracity (Vosoughi et al., 2018). Compared to verified information, they tend to be more emotionally intense and negatively toned (Liu et al., 2024). In contrast, institutional information is often less emotionally charged than misleading messages, making them less noticeable in attention-driven environments (Scheufele, 2014). Although emotional resonance explains engagement, research has focused more on diffusion and belief than on how emotional framing affects the structural resilience of information across contexts.

Misinformation makes use of aesthetics and storytelling techniques to increase its accessibility and immediacy. Media forms such as memes, videos, and infographics enable rapid comprehension and reduce perceived complexity, aligning with social media dynamics that favour easily consumable and attention-grabbing content (Tufekci, 2017).

Digital platforms enable both sophisticated manipulation and widespread user-generated visual content, contributing to the circulation of compelling misinformation (Weikmann & Lecheler, 2022). Although recent research has increasingly examined the role of visual design and aesthetics in credibility perception, much of the broader credibility literature remains focused on technical and informational features (Billard & Moran, 2022).

Institutional communication may rely on technical and nuanced language, which can reduce accessibility when not adequately adapted to the target audience (Fischhoff, 2013). As a result, misinformation often appears more direct and relatable. Although there have been several studies on how such characteristics influence engagement, fewer studies have explored their role in semantic stability.

2.2. The Role of Algorithms and Platform Logic

Although emotions and aesthetics make misinformation convincing, its far-reaching consequences rely more on the structural nature of online platforms. While misinformation occurs in all forms of media, it is especially common in digital spaces where visibility is driven by engagement metrics (Muhammed & Mathew, 2022).

In such settings, judgments about information reliability are often shaped by heuristic-based cues, including interface features, interactivity, and source indications, rather than an analytical approach (Sundar, 2008). Social media algorithms favour engagement-intensive content, which includes sensational, emotionally charged, or polarizing content (Cinelli et al., 2020). This may provide misinformation with a structural advantage in terms of diffusion and reach (Pariser, 2011).

Other elements like the presence of deceptive content, low user awareness, social bots, and changing platform dynamics further facilitate the dissemination of false information (Aïmeur et al., 2023), while personal factors including confirmation bias, trust perceptions, anxiety, and digital literacy affect the interpretation and sharing process of information by users (Muhammed & Mathew, 2022).

A recent study further concluded that young people and adults alike present limitations on the identification and verification of fake news, showing that frequent use of digital platforms does not guarantee media literacy competencies or resistance to misinformation (Trninić et al., 2022).

By contrast, high-quality information may receive less visibility if it lacks engagement attributes. In such scenarios, resonance often takes precedence over reliability in the dissemination of information (Tufekci, 2017). Although current literature explains amplification mechanisms, there is limited understanding of how platform dynamics interact with communicative characteristics to influence resistance to misinformation.

2.3. AI as a Double-Edged Sword on Misinformation

AI technologies have made misinformation more prolific and effective. According to a large-scale study conducted on 91,452 flagged posts by X, AI-driven misinformation was found to spread more widely than other content and, despite its lower credibility, still manage to go viral (Drolsbach & Pröllochs, 2025). In this context, AI-generated content can contribute to both misinformation and disinformation, further blurring boundaries, particularly when intent is unclear.

The generation of text, voice, and visual content that increasingly resembles human-produced communication (Monteith et al., 2024) has made it possible to produce large volumes of realistic yet misleading information (Simon et al., 2023), thus increasing the scale and complexity of the content that can be produced. This includes not only increased volume, but also greater sophistication and personalization of false content.

Advanced synthetic media, such as deepfakes, intensify this challenge. Human detection of deepfakes remains near chance levels (Diel et al., 2024), while increasingly realistic and emotionally engaging synthetic content may enhance the potential for manipulation and misinformation (Misirlis & Munawar, 2023). AI-generated explanations may further reinforce belief in false claims, even among analytically skilled users (Danry et al., 2025).

However, AI helps in mitigating these issues as well. For example, AI-based approaches can be used to boost media literacy and detect manipulation of content. Moreover, development of explainable AI, as well as natural language processing technologies, helps identify attempts at deceiving people (Zhou & Zafarani, 2020; Vishnupriya et al., 2024).

These insights imply that AI not only facilitates spreading of misinformation but also highlights the importance of evaluative approaches that include manipulability, contextual integrity and audience susceptibility beyond traditional accuracy-based criteria.

2.4. From Information Breakdown to Rethinking Quality

Over the past decade, multiple approaches have been proposed to improve the quality and trustworthiness of information in digital environments. Information-credibility models primarily examine source and message believability (Hovland et al., 1953; Sundar, 2008; Bates et al., 2006), while research on information resilience has focused on maintaining IQ by reducing vulnerabilities throughout the information lifecycle, particularly in organizational and digital information management contexts (Blay et al., 2020). Information integrity refers to the accuracy, consistency, and reliability of information across digital information ecosystems, particularly in response to misinformation and disinformation (United Nations, 2023). Information and media literacy initiatives have emphasized developing individuals' abilities to discover, critically evaluate, interpret, and recognize misinformation in digital environments (ACRL, 2016; Herrero-Diz & López-Rufino, 2021). Similarly, fact-checking approaches focus on verifying factual claims after publication (Nyhan et al., 2019). Although these perspectives contribute to improving IQ and trustworthiness from different angles, they provide limited support for systematically assessing the resilience of the information itself against manipulation, contextual distortion, emotional amplification, and platform-mediated dissemination. The RIQ framework was therefore developed to complement these existing perspectives by providing an integrated resilience-oriented assessment of information.

The limitations of traditional IQ models become more evident in contemporary information environments, as they often fail to account for the real-world conditions in which information is consumed today. With segmented audiences, algorithmic recommendations, and cognitive exhaustion, users often encounter content with limited contextual cues regarding provenance or communicative purpose. In addition, classic models typically view

quality as static and document-oriented, while misinformation today is fluid, fragmented and redeployed in different situations. In this context, what matters is the robustness of information to misinterpretation and emotionally charged frames, areas less studied by older models.

These trends illustrate the limitations of classic models of communication: the assumption of accuracy and completeness as primary indicators of quality. Instead, effective communication should include emotional awareness, visual accessibility, contextual relevance, and logic of digital media systems. Misinformation is effective not merely because it distorts the truth, but because it often aligns more closely with the communicative dynamics that drive engagement and belief formation (Lewandowsky et al., 2017). For this reason, standards used to assess the truthfulness of information will have to adapt accordingly, reflecting the realities of virality and algorithm-driven dissemination. In the current era, quality must increasingly be understood as extending beyond accuracy to include resistance to distortion, communicative transparency, and the interaction between the message and the audience and information environments that are susceptible to manipulation (Zrnec et al., 2022; López et al., 2024).

This shift calls for a reconceptualization of IQ as a dynamic, context-sensitive, and vulnerability-aware construct, one that accounts not only for content accuracy, but also for how information is shaped, interpreted, and transformed within complex communication environments.

3. Materials and Methods

This section presents the RIQ framework as a conceptual and analytical tool to assess the resilience of information against emotional manipulation, distortion, and misinterpretation.

The framework was created using a multi-phase process that involved: (1) recognizing the limitations of conventional IQ models; (2) abstracting key features from communication, cognition, and misinformation research; (3) iterative refinement through application. This ensured both intellectual grounding and internal coherence.

3.1. Research Design

The study adopts an exploratory research design aimed at developing and analytically applying RIQ. The approach combines conceptual development with a structured qualitative assessment to examine how information performs under complex communicative conditions.

The study focuses on analytical consistency and controlled variation, using adapted cases grounded in real-world communication patterns to evaluate the applicability and coherence of the proposed framework in different communication contexts.

3.2. Sample Selection and Analytical Approach

Samples were chosen using purposive sampling to ensure variation across critical dimensions such as emotional intensity, source transparency, communication purpose, and platform context. This approach enables the framework to be applied in a variety of contexts, allowing a more thorough analysis of how the dimensions behave across different message types.

The sample set consists of cases derived from recurring real-world communication patterns. These were developed through abstraction and synthesis of commonly observed forms of communication across multiple contexts, including misinformation, institutional messaging, and digital platform content, adapted for systematic evaluation. The design follows an analytical generalization approach to capture typical and recognizable forms of communication that reflect how information is commonly produced and shared.

The selection process sought to ensure variation in thematic content (such as climate change, health, public safety and finance) and in format (short posts, institutional announcements, phishing), platforms (from institutional communication channels to low-moderation social media environments), and emotional tones (ranging from fear-driven and alarmist to neutral and constructive).

The sample set includes:

- Examples adapted from real-world misleading content, designed to replicate persuasive communication patterns (e.g., climate misinformation claims).
- Representative samples that reflect recurring patterns of misinformation (e.g., conspiracy narratives, crisis alerts, scam messages).
- Adapted examples based on institutional or public communication formats (e.g., policy statements, public health guidance).

This selection supports both vertical (in-depth) and horizontal (comparative) analysis across dimensions, forming the analytical basis for the scoring and visualization procedures.

Given the exploratory nature of the study, interpretive judgment is treated as a structured analytical component, supported by a standardized rubric and inter-rater agreement procedures.

A detailed description of all samples, including their formulation and contextual rationale, is provided in Appendix B.

3.3. Conceptual Structure of the RIQ Framework

The framework draws on theories related to communication, cognitive psychology, and IQ.

The six dimensions are organized in two categories, as shown in Figure 1. The message's inherent qualities—manipulability resilience, interpretability, contextual transparency, and emotional framing—are addressed by information-centric dimensions. Audience resilience and platform suitability are two environment-centric dimensions that describe how information interacts with external circumstances.

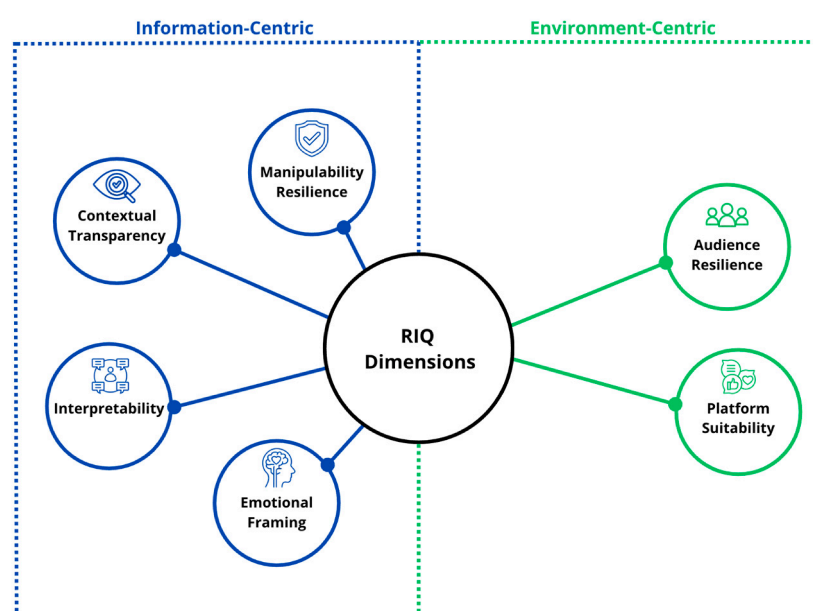


Figure 1. Information-centric and environment-centric dimensions. Source: Developed by the authors.

The selection of these dimensions followed an iterative conceptual synthesis. Initially, the literature review identified a broad set of constructs associated with IQ, credibility, misinformation, cognition, communication and digital platforms. Conceptually related

constructs were subsequently grouped into broader categories, while overlapping concepts were consolidated to maximize conceptual distinctiveness. Dimensions were retained when they captured a distinct aspect of information resilience that was not sufficiently represented by the remaining constructs, while redundant or highly overlapping concepts were integrated into broader dimensions. Traditional IQ attributes primarily associated with information management (e.g., timeliness or accessibility) were considered but were not retained as independent dimensions. This decision reflects the objective of the RIQ framework, which is to evaluate resilience against manipulation, reinterpretation, and platform-mediated dissemination rather than general information management quality.

Every dimension is grounded in well-established literature. Concerns with distortion and reframing in misinformation studies are reflected in manipulability resilience (Lewandowsky et al., 2017; Wardle & Derakhshan, 2017). The affect heuristic and dual-process theories are used in emotional framing (Kahneman, 2011; Slovic et al., 2007). Source credibility and mediated trust are related to contextual transparency (Hovland et al., 1953; Sundar, 2008). Cognitive load theory informs interpretability, particularly in relation to the processing demands imposed by complex information (Sweller, 1988). Research on misinformation susceptibility and analytical thinking is the foundation for audience resilience (Pennycook & Rand, 2019). Work on networked publics and platform logics is reflected in platform suitability (Boyd, 2010; van Dijck, 2013; Gillespie, 2014).

Though different in nature, the dimensions are inter-related. For example, low interpretability can lead to greater manipulability, whereas intense emotional framing can make the audience less resilient by fostering more intuition rather than analysis. Similarly, lack of contextual grounding can increase manipulability, particularly when information circulates online and it becomes detached from its original content. Information resilience therefore emerges not from individual dimensions in isolation, but from their combined configuration within specific communication environments.

3.4. Operationalization of RIQ Dimensions

Each of the six RIQ dimensions is measured using specific sub-criteria to facilitate a systematic assessment of information resilience in different scenarios.

3.4.1. Manipulability Resilience

Manipulability Resilience assesses how easily content can be distorted, reframed, or selectively interpreted in ways that alter its intended meaning.

This dimension is composed of three sub-criteria (Table 1):

Table 1. Manipulability Resilience sub-criteria.

Sub-Criterion	Description	Purpose
Ambiguity Resistance	Degree to which the content avoids ambiguous or open-ended wording.	Restricted interpretation minimizes the chances of misrepresentation or manipulation
Detail Specificity	Existence of specific details, sources or clarity in the message	Specific information is easier to confirm, decreasing possibilities of distortion or misunderstanding
Context Protection	Degree to which the message retains intended meaning when decontextualized	Well-anchored content is harder to twist or misuse outside its original setting

These three sub-criteria help assess the potential of information to be modified and weaponized even if the factual content is correct.

3.4.2. Contextual Transparency

Contextual Transparency evaluates the extent to which the message successfully conveys information about its origin, timeframe, and scope, allowing readers to judge its credibility and prevent misunderstanding. It comprises two sub-criteria (Table 2).

Table 2. Contextual Transparency sub-criteria.

Sub-Criterion	Description	Purpose
Source Transparency and Credibility	Clarity and specificity with which the message identifies its origin (e.g., author, institution, or source of evidence), enabling verification and assessment of trustworthiness.	Clear and explicit sourcing allows audiences to trace the information and evaluate its credibility, while vague or absent sourcing increases uncertainty and susceptibility to misinterpretation.
Time Reference	Whether the message includes a specific time frame or date relevant to the content.	Defining a specific time reference prevents misinterpretation by anchoring claims to a point in time.

Source, date and attribution can drastically affect how information is perceived and trusted. High contextual transparency allows audiences to assess credibility more effectively, especially in complex digital environments.

3.4.3. Interpretability

Interpretability captures the linguistic and structural accessibility of information, focusing on clarity, coherence, and suitability for diverse audiences and contexts. Its assessment relies on four sub-criteria (Table 3):

Table 3. Interpretability sub-criteria.

Sub-Criterion	Description	Purpose
Clarity	How easy the message is to understand at a surface level	Clear wording reduces confusion and helps readers grasp the main point quickly
Logical Flow	Whether ideas are connected in a coherent, step-by-step progression	Logical sequencing makes it easier for readers to follow the reasoning and absorb the message
Jargon Level	The degree to which the message uses complex, technical or domain-specific terms	High jargon can exclude non-experts or cause misunderstandings; low jargon improves accessibility
Length Appropriateness	Whether the length fits the task and audience	Optimal length ensures the message is neither vague nor exhausting to read

These aspects determine how effectively the information will be transferred across contexts or platforms, as they represent simplicity and clarity of both language and ideas, important aspects in ensuring that information is accurate and can be easily understood.

3.4.4. Emotional Framing

Emotional Framing is about how affective aspects are embedded in the content and how emotions influence perception, engagement and potential bias. It is important to note that the presence of emotion is not inherently negative, as it may also make the information more interesting, more engaging, or more compelling. However, when emotion overrides critical thinking or amplifies bias, it risks creating what has been called “emotional truth”, information that feels credible because it resonates emotionally, even if it is misleading.

This dimension is operationalized through two sub-criteria (Table 4):

Table 4. Emotional Framing sub-criteria.

Sub-Criterion	Description	Purpose
Emotional Restraint	Strength and vividness of the affective tone embedded in the message	High intensity increases salience and shareability but may reduce critical scrutiny
Moral Balance	Extent to which the message appeals to moral or ethical judgements (e.g., justice, outrage, duty, compassion)	Strong moral framing mobilizes action but may polarize interpretation and reduce nuance

Such features capture how emotional intensity and moral framing shape the tone and impact of a message, influencing not only whether it is noticed but also how it is processed and remembered.

3.4.5. Audience Resilience

Audience Resilience describes the audience’s ability to assess information without being misled, considering their emotional state, cognitive biases, and level of familiarity with digital technologies. Neglecting it could result in miscommunication, information overload, and manipulation. Such vulnerabilities may emerge among audiences with low institutional trust, heightened emotional exposure, or limited digital literacy.

It is evaluated using three sub-criteria (Table 5):

Table 5. Audience Resilience sub-criteria.

Sub-Criterion	Description	Purpose
Digital Literacy	Degree to which understanding the message requires prior knowledge, technical expertise, or advanced media literacy	High demand reduces accessibility and increases the risk of misinterpretation among general audiences
Emotional Safety	Extent to which the message avoids exploiting fear, empathy, outrage, or identity to trigger responses	Emotionally safe framing supports critical reflection, while strong exploitation may bypass it and disproportionately affect vulnerable groups
Inclusiveness of Appeal	Degree to which the message resonates broadly across audiences versus targeting niche or echo-chamber groups.	Broad appeal increases resilience, while selective targeting intensifies polarization and reduces interpretive balance

These criteria serve as indicators of how the features of the target audience influence the ability of a message to enlighten, persuade, or deceive. Audience resilience depends not only on the demographics and cognition of the group but also on their prior knowledge and media literacy.

3.4.6. Platform Suitability

Platform Suitability assesses how well content aligns with platform affordances, including format, shareability, and moderation dynamics, and how these influence its potential for amplification or distortion.

This dimension is assessed through three sub-criteria (Table 6):

Table 6. Platform Suitability sub-criteria.

Sub-Criterion	Description	Purpose
Format Alignment	Match between the message's length, structure, and style with the norms of the platform where it circulates	Strong alignment enhances visibility, clarity, and engagement within platform constraints; misalignment may reduce interpretability or distort the intended message
Safe Shareability	Likelihood of the message spreading without harm and within the platform's affordances (e.g., brevity, visuality, algorithmic boosts)	High engagement increases reach, which can amplify both credible and misleading content
Moderation Compliance	Probability that the message aligns with platform policies and avoids being flagged, challenged, or removed by moderation systems	High compliance ensures stable visibility and fosters audience trust, while low compliance reduces exposure and may prevent harmful amplification

Such elements capture whether a message is appropriately aligned with the characteristics and risks of the platform where it circulates. Platform suitability extends beyond form and length, requiring alignment with the dynamics of rapid dissemination.

3.5. Scoring Procedure

An explicit rubric was used to assess each of the six dimensions, ensuring consistency across evaluators. A five-point ordinal scale was adopted to provide sufficient discrimination between different levels of information resilience while maintaining interpretability and consistency across dimensions. Equal weighting was assigned to all dimensions because the framework was conceived as an exploratory multidimensional assessment and no theoretical or empirical evidence currently supports differential weighting.

From very low resilience (1) to very high resilience (5), the six dimensions, sub-criteria, and accompanying anchors in resilience are shown in Table 7.

These anchors provide a standard by which the ratings can be evaluated. The complete rubric is provided in Appendix A, Table A1.

Table 7. Dimensions, sub-criteria and interpretation scores.

Dimension	Sub-Criteria	1 (Very Low)	5 (Very High)
Manipulability Resilience	Ambiguity Resistance, detail specificity, context protection	Highly emotional, vague, easy to twist, manipulative authority	Neutral tone, precise and specific, hard to distort, transparent sources
Contextual Transparency	Source transparency and credibility, time reference	No source/date, vague, contextless	Fully sourced, dated, detailed
Interpretability	Clarity, logical flow, jargon level, length appropriateness	Dense, confusing, jargon-heavy, no examples	Clear, logical, jargon-free, well-illustrated
Emotional Framing	Emotional restraint, moral balance	Alarmist, moralizing	Neutral, balanced
Audience Resilience	Digital literacy, emotional safety, inclusiveness of appeal	Requires expertise, exploits fear, polarizing	Accessible to all, emotionally safe, broadly inclusive
Platform Suitability	Format alignment, safe shareability, moderation compliance	Viral but harmful, high flag risk	Well-adapted, shareable without harm, fully compliant

3.6. Inter-Rater Reliability

The scoring process was conducted by two independent evaluators using the structured rubric. Each sample was assessed independently across all dimensions, after which scoring outcomes were compared and discrepancies were resolved through discussion.

Inter-rater reliability was evaluated through Cohen's weighted kappa using quadratic weights (Cohen, 1968). Agreement levels were interpreted following the commonly used scale proposed in the literature (Landis & Koch, 1977).

The results are shown in Table 8.

Table 8. Cohen's weighted kappa across dimensions.

Dimension	Kappa (κ)	Interpretation
Manipulability Resilience	0.904	Almost perfect agreement
Contextual Transparency	0.783	Substantial agreement
Interpretability	0.785	Substantial agreement
Emotional Framing	0.840	Almost perfect agreement
Audience Resilience	0.896	Almost perfect agreement
Platform Suitability	0.966	Almost perfect agreement
Overall	≈ 0.908	Almost perfect agreement

The results indicate substantial agreement in the evaluation of all dimensions ($\kappa = 0.78\text{--}0.97$). The level of agreement overall was near perfect with $\kappa \approx 0.91$, indicating strong inter-rater reliability. Percentage agreement was lower, at approximately 62%, which is expected in ordinal scales with fine-grained distinctions. It is important to note that this statistic measures perfect agreement, and even minor differences (for example, by one point) are regarded as disagreement. In contrast, the Kappa coefficient offers a more robust measure of inter-rater agreement, as it accounts for partial agreement between raters, particularly in ordinal rating scales.

3.7. Classification Framework

Two complementary classification layers are defined: (i) a dimension-level categorization, where mean scores are mapped into performance levels (high, moderate, low), and (ii) a sample-level classification, where aggregated dimension scores are used to determine the overall structural robustness of the samples (critical, vulnerable, robust).

3.7.1. Performance-Level Categorization

To support interpretative analysis, mean dimension scores were categorized into performance levels. Table 9 presents the performance-level classification.

Table 9. Performance level classification based on mean scores.

Dimension Score Range	Performance Level
≥ 4	High
3–3.99	Moderate
< 3	Low

This categorization provides an interpretable layer for assessing the relative strength of each dimension. Although defined at the dimension level, this performance-level categorization is also applied to aggregated sample-level scores to provide a consistent interpretative scale across levels.

3.7.2. Structural Classification and Threshold Definition

In addition to continuous scoring, a threshold-based classification was introduced to capture structural vulnerabilities. Let n_1 , n_2 , and n_3 denote the number of dimensions falling within predefined score intervals corresponding to critical failures, moderate weaknesses, and borderline performance, respectively.

Table 10 specifies the interval-based thresholds used to transform mean dimensions scores into discrete categories (n_1 , n_2 , n_3), forming the basis for risk-based classification, with x being the dimension score.

Table 10. Interval-based thresholds for n .

n	Condition
n_1	$(x < 1.5)$
n_2	$(1.5 \leq x < 2.5)$
n_3	$(2.5 \leq x < 3.5)$

Based on the distribution of n_1 , n_2 , and n_3 , each sample is assigned to a structural classification category according to predefined rules. These rules are summarized in Table 11.

Table 11. Structural Classification rules.

Classification	Condition
Critical	$n_1 \geq 1$
Vulnerable	$n_2 \geq 2$ and $n_1 = 0$
Robust	$n_1 = 0$ and $n_2 \leq 1$

A single critical failure ($n_1 \geq 1$) is sufficient to classify a sample as critical, reflecting the assumption that one severe vulnerability may compromise overall information resilience. In the absence of critical failures, the vulnerable category captures the cumulative effect of multiple moderate weaknesses ($n_2 \geq 2$), indicating structural fragility despite the absence

of major failures. Scores falling within the n_3 interval are treated as borderline indicators and are used to support interpretative analysis rather than classification decisions. The proposed performance thresholds and structural classification rules were defined as heuristic interpretative categories intended to distinguish between predominantly weak, mixed, and consistently resilient information configurations. They should therefore be regarded as an initial conceptual specification for exploratory application and may be refined through future empirical validation. For ease of reference, the performance levels and structural classification rules are summarized in Appendix C (Tables A2 and A3).

4. Results

This section applies the RIQ Framework to the set of defined samples to examine how information resilience varies across different communicative conditions.

4.1. Case-Based Analysis

RIQ was applied to all fourteen samples. To avoid redundancy, two samples (4 and 9) were selected for detailed analysis, while the remaining samples are examined through comparative analysis and are described in Appendix B. The selection of contrasting cases aims to maximize analytical clarity, illustrating how the framework behaves under structurally divergent conditions.

The following case represents a high-resilience communication pattern:

Sample 4—Intergovernmental Panel on Climate Change (IPCC) AR6 Statement

“According to the Intergovernmental Panel on Climate Change’s sixth assessment report (AR6), limiting global warming to 1.5 °C requires rapid, deep and immediate cuts in greenhouse gas emissions across all sectors.” (Adapted from the IPCC Sixth Assessment Report (AR6) (IPCC, 2023); representative of institutional, evidence-based communication.)

Table 12 presents the dimensional assessment of Sample 4:

Table 12. Sample 4 analysis.

Dimension	Sub-Criteria Scores	Mean	Performance Level
Manipulability	5, 4.5, 5	4.83	High
Resilience			
Contextual	5, 4	4.50	High
Transparency			
Interpretability	4.5, 5, 3, 4.5	4.25	High
Emotional Framing	4.5, 4.5	4.50	High
Audience			
Resilience	3, 4.5, 4.5	4.00	High
Platform Suitability	4.5, 5, 5	4.83	High

With strong ratings across the board, this message presents a profile of structural robustness. By employing clear language and appropriate referencing, high resistance to manipulation and strong contextual clarity are achieved (IPCC AR6), reducing the risk of distortion and misinterpretation. The use of alarmist or moralized language is avoided, and emotional framing remains controlled with only a moderate level of urgency. Despite using domain-specific language, the message is easy to understand and maintains a balanced and non-manipulative emotional tone. Additionally, platform suitability scores are high, illustrating that the message is both engaging and safe. Strong structural robustness was shown by the absence of critical failures ($n_1 = 0$; $n_2 = 0$), consistent with this profile.

In contrast, the following sample represents a low-resilience communication pattern:

Sample 9—Contaminated Water Alert

“URGENT: Tap water has been contaminated. Authorities are hiding the truth. Do NOT drink it!” (Representative of false public safety alerts documented during crisis events and debunked by Snopes)

Table 13 presents the dimensional assessment of Sample 9:

Table 13. Sample 9 analysis.

Dimension	Sub-Criteria Scores	Mean	Performance Level
Manipulability	1, 1, 1	1.00	Low
Resilience			
Contextual	1, 1.5	1.25	Low
Transparency			
Interpretability	4, 3, 5, 3.5	3.88	Moderate
Emotional Framing	1.5, 1	1.25	Low
Audience			
Resilience	5, 1, 4	3.33	Moderate
Platform Suitability	5, 1, 1	2.33	Low

This message shows consistently low resilience across all dimensions. The message provides little protection against manipulation and lacks contextual grounding; moreover, it relies extensively on non-specific claims and does not provide references, making it easily susceptible to distortion and misinterpretation. Emotional framing is dominant and unrestrained, characterized by alarmist tone, accusatory language, and explicit appeals to fear and urgency. Though the message has a relatively high level of interpretability because of its clear and easy language, it lacks coherent reasoning, with assumptions substituting the facts and logical arguments. This accessibility makes it highly vulnerable to manipulation, especially considering its low level of emotional safety, particularly among distrust-prone audiences. Platform suitability presents a paradox: while highly optimized for visibility and engagement, its shareability is unsafe and likely to trigger platform moderation due to misleading and alarmist content. Critical failures ($n_1 = 3$) in multiple categories show systemic structural vulnerability. The example illustrates how platform conformance, emotional impact, and simplicity may be used to promote virality while significantly undermining the legitimacy and strength of the content.

4.2. Comparative Analysis Across Samples

Based on the individual analysis presented earlier, a comparative approach was adopted to examine how resilience profiles vary across samples, highlighting structural differences and recurring patterns across the dataset.

Figure 2 presents a comparison between two contrasting samples to highlight differences in resilience profiles. While based on selected cases, this visualization reflects broader patterns observed across the full set of analysed messages.

The radar chart illustrates the sharp contrast between samples 4 and 9. The former exhibits strong transparency, resilience, and credible sourcing, while the latter exhibits significant weaknesses in transparency and resilience but scores moderately high in interpretability and audience resilience.

Extending this comparison across all fourteen samples, the heatmap below (Figure 3) enables comparison across the six dimensions of the framework, highlighting variations between content types. By visualizing these dimensions simultaneously, the heatmap facilitates the identification of recurring patterns and contrasts across samples.

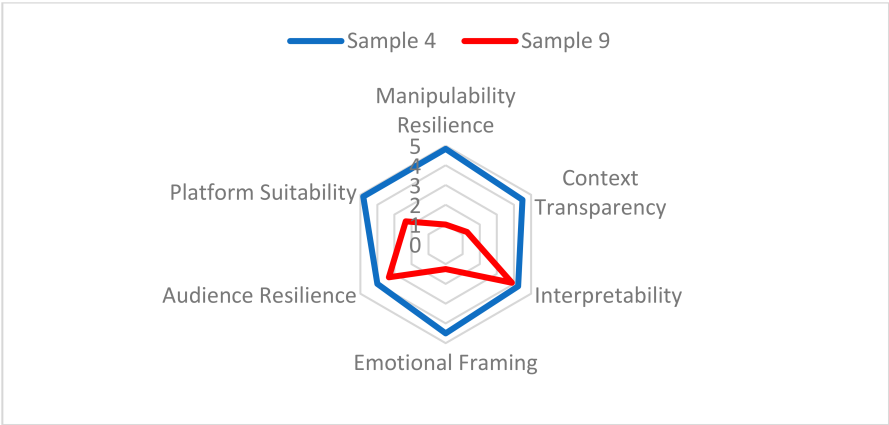


Figure 2. Dimensions Average Scores (generated in Excel).

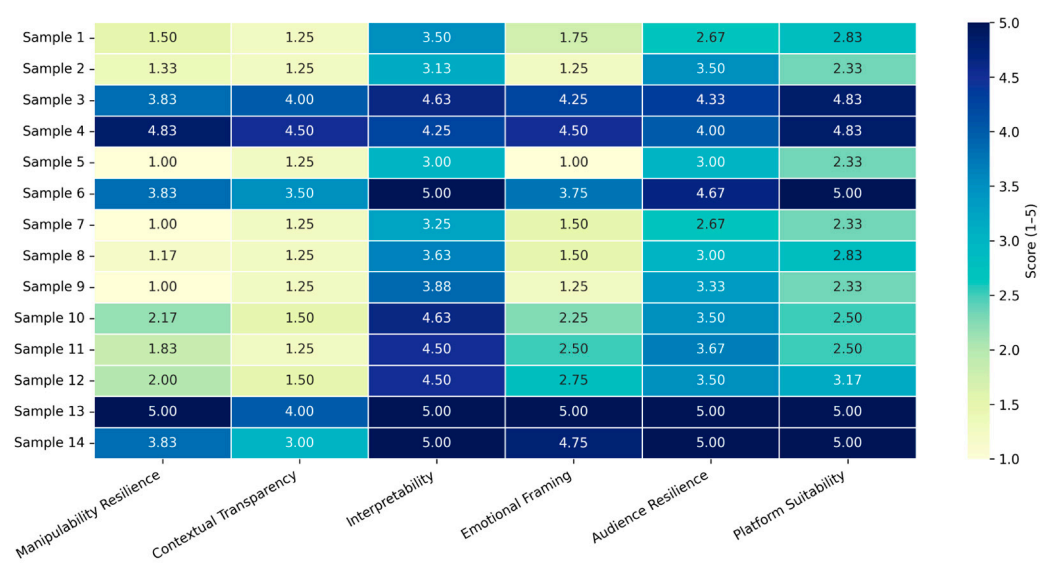


Figure 3. IQ dimensions across samples (generated in Python 3.12.13).

According to the heatmap, content with poor resilience (Samples 1, 2, 5, 7, 8, and 9) exhibits low scores in Manipulability Resilience, Contextual Transparency, and Emotional Framing. This shows that the content is more prone to distortion, has fewer sources, and exhibits problematic emotional framing. High-resilience content, on the other hand, has strong and consistent scores in all aspects, particularly in transparency and structural coherence (Samples 3, 4, 13, and 14). The majority of samples, even those with lower-quality content, have comparatively high interpretability, indicating that information reliability is not always correlated with clarity and accessibility. Samples 6, 10, 11, and 12 are examples of intermediate situations that exhibit more heterogeneous profiles, combining different levels of manipulability resilience with moderate transparency and emotional control.

4.3. Integrated Assessment

In addition to the dimensional assessment, a global indicator was derived from the radar chart by calculating the area of the polygon formed by the six dimensions, assuming equal angular spacing. Table 14 outlines the integrated assessment structure of the RIQ framework, combining continuous scoring, performance categorization, threshold-based classification, and geometric aggregation.

Table 14. Integrated assessment structure of the RIQ framework.

Level	Metric	Definition	Formalization	Interpretation
Sub-criteria	Raw Score	Individual evaluation	1–5 scale	Base measurement
Dimension/Sample (continuous)	Mean Score ($\bar{x}$)	Average across sub-criteria	$\bar{x} = \frac{1}{n} \sum x_i$	Overall performance level
	Performance Level	Categorization of mean scores	Threshold-based (≥ 4 high, 3–3.99 moderate, <3 low)	Overall effectiveness
Sample (risk-based)	Critical	≥ 1 critical failure	$n_1 \geq 1$	Single-point failure
	Vulnerable	Accumulated weaknesses	$n_1 = 0$ and $n_2 \geq 2$	Cumulative fragility
	Robust	No significant weaknesses	$n_1 = 0$ and $n_2 \leq 1$	Structurally stable
	(Flag) Borderline	High concentration of mid-range scores	$n_3 \geq 2$	Interpretative signal (no classifying)
Global (geometric)	Radar Area (A)	Polygon area from all dimensions	$A = \frac{1}{2} \sin\left(\frac{2\pi}{d}\right) \sum_{i=1}^d r_i r_{i+1} *$	Overall multidimensional performance

* With $r_{n+1} = r_1$, where r_i represents the mean score of each dimension, corresponding to the radial distance in the radar chart, and d the number of dimensions (6).

The radar area gives a numerical measure of overall performance in addition to the visual comparison. Sample 4 has a much larger area than Sample 9, indicating a higher level of multidimensional IQ. Although radar area is positively associated with overall performance, it is not determined solely by mean scores. Instead, it reflects both the magnitude and the distribution of values across dimensions, meaning that structurally imbalanced profiles may reflect different area values despite similar averages.

Table 15 presents the integrated assessment of all samples. It combines the overall mean scores for performance, the n -values for structural classification based on threshold conditions, and the radar area as a global geometric indicator. This table illustrates how overall IQ comes from the interaction between performance and structural robustness across samples.

Table 15. Integrated assessment of all samples.

Sample	Overall Mean Score	Performance Level	Structural Condition (n)	Structural Classification (n)	Radar Area	Overall Assessment
1	2.25	Low	$n_1 \geq 1$	Critical	12.43	Unreliable
2	2.13	Low	$n_1 \geq 1$	Critical	10.95	Unreliable
3	4.31	High	$n_1 = 0$ and $n_2 \leq 1$	Robust	48.15	Trustworthy
4	4.49	High	$n_1 = 0$ and $n_2 \leq 1$	Robust	52.23	Trustworthy
5	1.93	Low	$n_1 \geq 1$	Critical	8.80	Unreliable
6	4.29	High	$n_1 = 0$ and $n_2 \leq 1$	Robust	47.58	Trustworthy
7	2.00	Low	$n_1 \geq 1$	Critical	9.48	Unreliable
8	2.23	Low	$n_1 \geq 1$	Critical	11.73	Unreliable
9	2.17	Low	$n_1 \geq 1$	Critical	10.63	Unreliable
10	2.76	Low	$n_1 = 0$ and $n_2 \geq 2$	Vulnerable	18.71	Weak
11	2.71	Low	$n_1 \geq 1$	Critical	17.74	Unreliable
12	2.90	Low	$n_1 = 0$ and $n_2 \geq 2$	Vulnerable	20.86	Weak
13	4.83	High	$n_1 = 0$ and $n_2 \leq 1$	Robust	60.62	Trustworthy
14	4.43	High	$n_1 = 0$ and $n_2 \leq 1$	Robust	50.37	Trustworthy

The integrated assessment, which integrates structural categorization (n -values) and performance level (mean scores), shows a distinct polarization throughout the examined samples. Low-performing samples (Samples 1, 2, 7, 8, 9 and 11) are classified as unreliable because they are consistently associated with critical structural problems ($n_1 \geq 1$). Ad-

ditionally, these examples have the lowest radar areas, which highlights their structural fragility and suggests poor multidimensional performance.

On the other hand, top-performing samples (Samples 3, 4, 6, 13, and 14) consistently exhibit high radar areas due to their strong mean scores and structural robustness ($n_1 = 0$ and $n_2 \leq 1$). Their designation as trustworthy is supported by this alignment across performance, structural stability, and geometric aggregation, indicating that high IQ is defined by both strong individual aspects and consistency across them.

Intermediate cases (Samples 10 and 12) exhibit a distinct pattern. Although their overall performance remains low, the absence of critical failures ($n_1 = 0$) combined with multiple moderate weaknesses ($n_2 \geq 2$) results in a vulnerable classification. These samples display higher radar areas than other low-performing cases, indicating a more balanced but still fragile structure. Their classification as weak reflects this intermediate position between structural failure and robustness.

Overall, the findings show that robust configurations typically result in consistently good performance across dimensions, while also highlighting a clear correlation between structural vulnerabilities and low performance. IQ in the analysed dataset reflects a clear dichotomy between structurally resilient and structurally fragile communication patterns, according to the limited number of intermediate profiles.

This analysis indicates that major vulnerabilities ($n_1 \geq 1$) are present in a number of samples with moderate or high interpretability when combined with threshold-based classification, suggesting that accessibility and clarity do not always equate to resilience. This difference highlights the importance of distinguishing between structural resilience and overall performance.

5. Discussion

This section discusses the findings derived from the proposed framework. As an exploratory application of the framework, the analysis illustrates how the RIQ dimensions can be operationalized across different communication contexts. It provides insight into how communication styles influence the way content is received and assessed in terms of credibility. The section also reflects its limitations and broader implications of the framework for future research.

5.1. Interpretation of Findings

Across the fourteen samples, three consistent patterns emerge. Emotionally charged messages typically exhibit weaker resistance to manipulation, lower contextual clarity and audience resilience. Despite being very appealing and widespread, these communications are structurally fragile and prone to manipulation, as they rely on ambiguity, alarmism, and lack of clear authority sourcing. Fraudulent messages reinforce this pattern by using emotional and moral appeals simultaneously.

By contrast, institutional and evidence-based messages exhibit high resilience due to strong sourcing, contextual grounding, and logical coherence. Yet, formal tone and technical jargon may limit accessibility and adaptability in fast-moving contexts. Positive emotions can coexist with a high degree of clarity and comprehension, promoting both engagement and informational integrity, as illustrated by constructive and campaign-based messaging.

Interpretability remains relatively high across most samples, including lower-resilience content, suggesting that readability alone is not a reliable indicator of IQ. Clear and accessible messages may still lack contextual grounding or exhibit high susceptibility to manipulation.

Platform suitability shows greater variation across samples. Although institutional and campaign-related messages usually align with platform norms without compromising their integrity, messages with low resilience are often optimized for visibility and engagement.

This suggests a structural tension in which emotionally activated and simplified content may achieve higher dissemination potential despite weaker IQ.

The integrated assessment reveals a clear polarization across samples, with low-performing cases consistently associated with critical structural vulnerabilities, and high-performing cases exhibiting both strong scores and structural robustness. Intermediate profiles remain limited, typically reflecting structurally vulnerable configurations despite moderate performance. Radar area values reflect these differences in overall performance, although not fully capturing internal structural imbalances.

The findings show a difference between overall performance and structural robustness that extends beyond average scores. Critical vulnerabilities ($n_1 \geq 1$) are present in several samples that exhibited moderate or even high scores in certain dimensions, suggesting that isolated flaws can undermine resilience even when overall performance appears otherwise satisfactory.

Overall, the patterns point to complex interactions between structural clarity, contextual anchoring, and emotional intensity in various communication contexts. The most engaging messages might not be the most durable, highlighting the significance of assessing IQ in addition to accuracy and considering how content functions in situations of amplification, reinterpretation, and platform-driven visibility.

5.2. Theoretical Implications

This framework advances literature by reframing IQ in terms of resilience rather than truth-value alone.

Beyond continuous scoring, the model also includes threshold-based classification to identify structural vulnerabilities, enabling a distinction between overall performance and systemic fragility. This allows for the detection of critical weaknesses that may not be visible through aggregate scores alone. More broadly, RIQ shifts the focus from evaluating whether information is correct to understanding how well it withstands distortion, reinterpretation, and amplification across dynamic communication environments.

While traditional fact-checking and accuracy-oriented models differentiate between truths and falsehoods, this model draws attention to the structures that make information either vulnerable or resilient to manipulations, decontextualization, and emotionally charged contexts. The proposed framework introduces six evaluative dimensions spanning both content-intrinsic and environment-centric factors. In doing so, the framework extends existing models of IQ by integrating resilience-oriented, contextual, and audience-related dimensions into a unified evaluative approach.

This shift in viewpoint moves the focus from the veracity of the information to how resistant it is to distortion, helping explain why false messages may achieve high levels of trust due to their emotional or structural appeal, while others, even if they are accurate, may lack credibility.

From a theoretical perspective, the framework contributes by bridging IQ research with established models in communication and cognition. First, with its integration of emotional framing and audience susceptibility, the model reflects the dual-process theory (Kahneman, 2011) and the affect heuristic principle (Slovic et al., 2007), acknowledging that credibility is shaped not only by informational attributes but also by cognitive and affective aspects of the information process. Additionally, contextual and platform transparency position the framework within theories of online credibility such as source credibility theory (Hovland et al., 1953) and the MAIN model (Sundar, 2008). In this sense, RIQ does not replace existing IQ frameworks but extends them into environments where information is dynamic, socially mediated, and vulnerable to manipulation.

From a methodological point of view, the framework is built on a combination of continuous scoring, classification by threshold, and geometrical aggregation. This makes it possible to

identify not only the extent of performance but also its vulnerability, addressing limitations of purely average-based approaches. The distinction between single-point failures and cumulative weaknesses introduces a resilience-oriented perspective, providing a structured mechanism for identifying vulnerabilities that remain invisible in traditional IQ models.

5.3. Practical Implications

In addition to its theoretical implications, the framework offers practical relevance for multiple stakeholders. It can be used by researchers as a method of analysis for comparative purposes, and by journalists and educators to improve communication strategies. Its criteria can be adapted by policymakers for designing resiliency-based measures, while technology developers can incorporate the framework into automated or semi-automated processes for detecting vulnerable or manipulative content. These stakeholder implications are presented in Table 16.

Table 16. Implications for stakeholders.

Stakeholder	How the RIQ Framework Can Be Used
Researchers	Utilize the six dimensions as a coding scheme in content analysis studies for comparing different samples and monitoring message resilience
Journalists & Communicators	Analyse and correct any weaknesses in their message design (such as lack of transparency or interpretability) to enhance resilience before publication
Educators & Media Literacy Programs	Teach students to identify emotional appeals, lack of context, and manipulation strategies by applying the framework as a checklist
Policy-Makers & Regulators	Inform policy guidelines and risk assessments by identifying message types that are susceptible to misinterpretation or harmful amplification
Platform Designers	Integrate the framework's dimensions into automated tools for flagging high-risk content or requesting context information before sharing

The framework can support information systems by enabling more context-aware and resilience-oriented evaluation of IQ in dynamic environments.

6. Conclusions

The motivation for this study arises from a key limitation in previous research on IQ. While traditional models of IQ have provided important criteria such as accuracy, completeness, timeliness, and relevance, they have not provided a comprehensive explanation of how information operates in today's digital environments, as these perspectives have not yet been sufficiently integrated into a comprehensive perspective to evaluate whether information is capable of preserving its clarity, credibility, and intended meaning in the face of distortion and reinterpretation.

In response to this gap, the study was guided by three research questions. In relation to RQ1, which sought to understand what dimensions are necessary to assess information resilience in addition to traditional IQ characteristics, the study proposed six complementary dimensions: Manipulability Resilience, Contextual Transparency, Interpretability, Emotional Framing, Audience Resilience, and Platform Suitability. These dimensions go beyond traditional IQ testing, as they account for not only the intrinsic properties of information, but also its susceptibility to distortion, its grounding in context, its affective framing, its risks to audiences, and its appropriateness for safe circulation in platform-mediated environments.

Regarding RQ2, which asked how these dimensions can be operationalized into a structured analytical framework, the study developed a scoring rubric based on specific sub-criteria for each dimension. The framework was applied through a structured qualitative assessment, supported by threshold-based classification, inter-rater reliability analysis, and complementary visual and geometric tools, including heatmaps, radar charts, and radar-area calculations. The inter-rater reliability results indicated substantial to almost perfect agreement across the framework dimensions, suggesting that the scoring procedure provides a consistent basis for exploratory application.

Regarding RQ3, which asked how the RIQ framework distinguishes between critical, vulnerable, and robust communication patterns, the exploratory findings suggest that information resilience cannot be inferred from clarity or accessibility alone. Several low-resilience samples were highly interpretable, easy to process, and technically adapted to platform dynamics, yet remained structurally fragile because they lacked contextual transparency, relied on emotional urgency, or were highly susceptible to manipulation. In contrast, institutional and evidence-based communication samples showed stronger resilience due to clearer sourcing, contextual grounding, logical coherence, and lower susceptibility to distortion.

The classification procedure illustrated three main information patterns within the analysed sample. First, critical or unreliable configurations were associated with low transparency, weak resistance to manipulation, alarmist emotional framing, and unsafe shareability. Second, vulnerable or intermediate configurations showed no single critical failure but included cumulative weaknesses across several dimensions, indicating that moderate performance does not necessarily ensure structural robustness. Third, robust or trustworthy configurations were characterized by consistent performance across dimensions, strong contextual anchoring, controlled emotional framing, and high resistance to distortion. These results suggest that RIQ may be useful not only for measuring average performance, but also for identifying structural vulnerabilities that may remain hidden in aggregate IQ scores.

The study therefore concludes that IQ should no longer be understood only as a property of accurate, complete, or timely information. In contemporary platform-mediated environments, high-quality information must also be resilient: it must be able to preserve meaning, credibility, and contextual integrity when exposed to emotional framing, rapid sharing, algorithmic amplification, and adversarial reinterpretation. The RIQ framework contributes to this reconceptualization by integrating insights from IQ research, misinformation studies, communication theory, cognitive psychology, platform studies, and media literacy into a single analytical model.

Nevertheless, several limitations should be acknowledged. The empirical component of this study was exploratory and based on a limited set of representative samples rather than a large-scale dataset. The cases were adapted from recurring real-world communication patterns, but they do not capture the full complexity, diversity, or temporal evolution of online information flows. In addition, although the scoring rubric and inter-rater reliability procedures improved analytical consistency, the evaluation still involved interpretive judgment. The threshold-based classification and radar area indicator should therefore be interpreted as exploratory analytical tools rather than definitive validation instruments. Finally, the framework was designed primarily for platform-mediated communication contexts and may require adaptation for highly regulated, technical, legal, scientific, or organizational information environments.

Although the illustrative communication samples illustrate the operational applicability of the RIQ framework, they should be considered as an exploratory application rather than a comprehensive empirical validation of the proposed model. Future research should address these limitations through broader empirical validation. Priority directions include large-scale dataset testing, cross-platform comparison, audience-response validation, automated RIQ assessment using NLP and AI-based tools, longitudinal tracking of information resilience,

sensitivity analysis of scoring thresholds, and sector-specific applications in areas such as public health, climate communication, finance, education, and institutional communication. Further studies should also analyse how RIQ dimensions interact with algorithmic amplification, user trust, media literacy, and behavioural responses. By advancing these directions, the RIQ framework can support more resilient communication practices, more context-aware information systems, and more effective strategies for countering misinformation in fragmented and emotionally driven digital environments.

Author Contributions: Conceptualization, A.R.A. and F.G.; Methodology, A.R.A. and F.G.; Formal analysis, A.R.A. and F.G.; Investigation, A.R.A. and J.B.; Writing—original draft, A.R.A.; Writing—review & editing, F.G. and J.B. Supervision, F.G. and J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This article was developed within the scope of the ADT4Blue project, under EAPA_0061/2022, co-funded by the FEDER-Interreg Atlantic Area.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are included in the article and its appendices. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Scoring Rubric

Table A1. Full scoring rubric.

Dimension	Sub-Criterion	1 (Very Low Resilience)	2 (Low)	3 (Moderate)	4 (High)	5 (Very High Resilience)
Manipulability Resilience	Ambiguity Resistance (Can this message be plausibly interpreted in multiple ways?)	No resistance; highly ambiguous wording enables multiple interpretations	Limited resistance; frequent ambiguity allows reinterpretation	Partial resistance; some ambiguity but meaning generally clear	Strong resistance; mostly explicit and constrained meaning	Full resistance; precise, unambiguous, and resistant to reinterpretation
	Detail Specificity (Does the message contain concrete, verifiable details that can be fact-checked?)	No verifiable details; entirely vague	Limited details; unclear or ambiguous references	Partial details; some verifiable elements but incomplete	Strong detail; clear and specific references	Full detail; highly precise, verifiable, and comprehensive
	Context Protection (Does the message retain its intended meaning when removed from its original context?)	Meaning collapses outside original context	Limited protection; easily distorted when decontextualized	Partial protection; retains some meaning outside context	Strong protection; meaning largely preserved when quoted	Full protection; meaning fully robust across contexts

Table A1. Cont.

Dimension	Sub-Criterion	1 (Very Low Resilience)	2 (Low)	3 (Moderate)	4 (High)	5 (Very High Resilience)
Contextual Transparency	Source Transparency and Credibility (Is the source of the information clearly identifiable and credible?)	No source identified	Limited source indication; vague or unclear reference	Partial transparency; source hinted but incomplete	Strong transparency; clearly identified and credible source	Full transparency; explicit, specific, and authoritative source
	Time Reference (Is the temporal context of the information clearly specified?)	No temporal reference	Limited reference; vague timing (e.g., “recently”)	Partial reference; general timeframe implied	Strong reference; clear but broad timeframe	Full reference; precise and explicit date/time
Interpretability	Clarity (Is the message immediately understandable to a general audience?)	No clarity; confusing or unintelligible wording	Limited clarity; meaning often unclear	Partial clarity; generally understandable with some ambiguity	Strong clarity; mostly clear and accessible	Full clarity; fully clear, precise, and easily understood
	Logical Flow (Are the ideas presented in a coherent and logically consistent manner?)	No logical structure; disconnected elements	Limited structure; weak or inconsistent connections	Partial structure; some logical progression	Strong structure; coherent and well-organized	Full structure; fully logical, consistent, and sequential
	Jargon Level (Does the message avoid unnecessary technical or specialized language?)	Excessive jargon; inaccessible to most audiences	High jargon; frequent barriers to understanding	Moderate jargon; partially accessible	Low jargon; mostly accessible	No unnecessary jargon; fully accessible to general audiences
	Length Appropriateness (Is the length of the message appropriate for conveying its content effectively?)	Inappropriate length; severely too short or too dense	Limited appropriateness; often imbalanced	Partial appropriateness; generally acceptable length	Strong appropriateness; well-balanced for purpose	Full appropriateness; optimal length for clarity and context
Emotional Framing	Emotional Restraint (Does the message avoid inducing strong emotional reactions such as fear, urgency, or outrage?)	No restraint; highly alarmist or emotionally intense	Limited restraint; strong emotional tone	Partial restraint; moderate emotional expression	Strong restraint; mostly neutral tone	Full restraint; consistently neutral and balanced
	Moral Balance (Does the message avoid polarizing or absolutist moral framing?)	No balance; extreme or absolutist moral framing	Limited balance; strongly polarizing framing	Partial balance; moderate moral positioning	Strong balance; mostly nuanced framing	Full balance; fully nuanced and non-polarizing

Table A1. Cont.

Dimension	Sub-Criterion	1 (Very Low Resilience)	2 (Low)	3 (Moderate)	4 (High)	5 (Very High Resilience)
Audience Resilience	Digital Literacy Demand (Can the message be accurately understood without requiring advanced knowledge or expertise?)	Requires expert knowledge	High literacy demand; difficult for general audiences	Moderate literacy demand; partially accessible	Low literacy demand; mostly accessible	Minimal literacy demand; fully accessible to general audiences
	Emotional Safety (Does the message avoid exploiting emotional vulnerabilities?)	No safety; exploits fear, outrage, or empathy	Limited safety; frequent emotional manipulation	Partial safety; moderate manipulation risk	Strong safety; mostly non-exploitative	Full safety; consistently non-exploitative and emotionally safe
	Inclusiveness of Appeal (Is the message accessible and relevant to a broad and diverse audience?)	No inclusiveness; restricted to niche or echo chamber	Limited inclusiveness; narrow audience appeal	Partial inclusiveness; moderate audience reach	Strong inclusiveness; broadly accessible	Full inclusiveness; universally accessible and inclusive
Platform Suitability	Format Alignment (Is the message well adapted to the conventions and format of the platform?)	No alignment; incompatible with platform norms	Limited alignment; weak fit to platform	Partial alignment; some adaptation	Strong alignment; mostly aligned with platform norms	Full alignment; fully optimized for platform
	Safe Shareability (Can the message be shared without causing harm or misinformation?)	Unsafe; likely to cause harm if shared	Limited safety; high risk of harm	Partial safety; moderate risk	Strong safety; low risk	Full safety; safe to share with minimal risk
	Moderation Compliance (Is the message likely to comply with platform moderation policies?)	Non-compliant; likely to be removed or flagged	Limited compliance; high risk of flagging	Partial compliance; possible risk	Strong compliance; low risk	Full compliance; fully aligned with platform policies

Appendix B. Samples

Each sample was constructed as a minimal, representative instance of a broader communication pattern, preserving key structural features while avoiding unnecessary contextual noise.

Sample 1—Climate volcano claim

“Claims circulating online suggest that volcanic CO₂ emissions exceed human emissions and therefore climate change is not driven by human activity.” (Adapted from misinformation claims fact-checked by NASA and NOAA)

Sample 2—Earthquake donation scam

“Urgent: Thousands still buried under rubble. You can help save a child’s life. Donate here now.” (Representative of fraudulent donation appeals reported)

after the 2023 Turkey–Syria earthquake, documented by Interpol and Federal Trade Commission.)

Sample 3—Warning on AI scams

“AI-generated scams are widespread and becoming increasingly difficult to detect. Scammers can use information available on the internet, including images and audio from social media, to convince people that the voice on the other end of the call is someone they can trust.” (Adapted from public warnings issued by the [State of California Department of Justice, 2024](#).)

Sample 4—Intergovernmental Panel on Climate Change (IPCC) AR6 Statement

“According to the Intergovernmental Panel on Climate Change’s sixth assessment report (AR6), limiting global warming to 1.5 °C requires rapid, deep and immediate cuts in greenhouse gas emissions across all sectors.” (Adapted from the IPCC Sixth Assessment Report (AR6) ([IPCC, 2023](#)); representative of institutional, evidence-based communication.)

Sample 5—Chemtrails

“They are spraying us again!!! Wake up before it is too late. Look at those lines—they are poisons, not contrails.” (Representative of “chemtrail” conspiracy content widely debunked by European Space Agency and NASA.)

Sample 6—Tree Campaign

“Thanks to your support, we just surpassed 50 million trees planted! Each sapling means hope, cleaner air and new jobs for rural families.” (Representative of milestone posts by Eden Reforestation Projects and similar NGOs; adapted for illustration.)

Sample 7—Vaccine Microchip

“Doctors won’t tell you this, but the new vaccines contain tracking microchips activated by 5G towers.” (Representative of COVID-19 misinformation narratives documented by World Health Organization and Reuters Fact Check)

Sample 8—Election Fraud

“Thousands of votes mysteriously disappeared overnight. Observers are being silenced.” (Representative of election misinformation narratives documented in reports by European Commission and Pew Research Centre)

Sample 9—Contaminated Water Alert

“URGENT: Tap water has been contaminated. Authorities are hiding the truth. Do NOT drink it!” (Representative of false public safety alerts documented during crisis events and debunked by Snopes)

Sample 10—Bank Phishing Message

“Your bank account has been temporarily locked due to suspicious activity. Please verify your identity immediately to avoid permanent suspension: [link]” (Adapted from phishing patterns reported by Europol and Federal Trade Commission)

Sample 11—Job Scam

“We are offering flexible remote positions with a weekly salary of €1,200. No experience required.” (Representative of online job scams documented by Europol and LinkedIn safety reports)

Sample 12—Health Misinformation

“Recent studies suggest that natural immunity provides stronger protection than vaccines in most cases. Experts are beginning to reconsider long-standing assumptions.” (Representative of health misinformation narratives documented dur-

ing COVID-19 and analysed by World Health Organization and Centres for Disease Control and Prevention)

Sample 13—Public Health Guidance

“Regular handwashing with soap significantly reduces the spread of infectious diseases. It is recommended to wash hands for at least 20 s, especially before eating and after public exposure.” (Adapted from standard public health communication guidelines issued by international organizations such as the World Health Organization.)

Sample 14—Mental Health Awareness Campaign

“It’s okay to not feel okay. Talking to someone you trust can make a difference. Small steps matter—reach out, listen, and support each other.” (Representative of mental health awareness campaigns disseminated by NGOs and public institutions.)

Appendix C. RIQ Classification Procedure

Table A2. Performance Levels.

Mean Score	Performance Level	Rationale
≥ 4	High	Strong and consistent resilience across the evaluated sub-criteria
3.00–3.99	Moderate	Acceptable performance with identifiable weaknesses
<3	Low	Insufficient resilience requiring attention

Table A3. Structural classification.

Classification	Rule	Conceptual Rationale
Critical	$n_1 \geq 1$	A single critical weakness may substantially compromise the overall resilience of the information
Vulnerable	$n_2 \geq 2$ and $n_1 = 0$	Although no critical failures exist, multiple moderate weaknesses may accumulate and reduce structural robustness
Robust	$n_1 = 0$ and $n_2 \leq 1$	The information demonstrates consistently strong performance with no significant vulnerabilities.

The borderline indicator ($n_3 \geq 2$) is used as an interpretative signal rather than an independent classification category, identifying cases with a concentration of intermediate scores that may require qualitative interpretation.

References

- Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. *Social Network Analysis and Mining*, 13, 30. [CrossRef] [PubMed]
- Association of College and Research Libraries (ACRL). (2016). *Framework for information literacy for higher education*. American Library Association. Available online: <https://www.ala.org/acrl/standards/ilframework> (accessed on 1 July 2026).
- Bates, B. R., Romina, S., Ahmed, R., & Hopson, D. (2006). The effect of source credibility on consumers’ perceptions of the quality of health information on the Internet. *Medical Informatics and the Internet in Medicine*, 31(1), 45–52. [CrossRef] [PubMed]
- Battineni, G., Baldoni, S., Chintalapudi, N., Sagaro, G. G., Pallotta, G., Nittari, G., & Amenta, F. (2020). Factors affecting the quality and reliability of online health information. *Digital Health*, 6, 1–11. [CrossRef] [PubMed]

- Billard, T. J., & Moran, R. E. (2022). Designing trust: Design style, political ideology, and trust in “fake” news websites. *Digital Journalism*, 11, 519–546. [\[CrossRef\]](#)
- Blay, K. B., Yeomans, S., Demian, P., & Murguia, D. (2020). The information resilience framework: Vulnerabilities, capabilities, and requirements. *Journal of Data and Information Quality*, 12(3), 1–25. [\[CrossRef\]](#)
- Boyd, D. (2010). Social network sites as networked publics: Affordances, dynamics, and implications. In Z. Papacharissi (Ed.), *A networked self* (pp. 47–66). Routledge. [\[CrossRef\]](#)
- Cinelli, M., Quattrocioni, W., Galeazzi, A., Valensise, C. M., Brugnoti, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports*, 10, 16598. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. [\[CrossRef\]](#) [\[PubMed\]](#)
- Danry, V., Pataranutaporn, P., Groh, M., & Epstein, Z. (2025). Deceptive explanations by large language models lead people to change their beliefs about misinformation more often than honest explanations. In *Proceedings of the 2025 CHI conference on human factors in computing systems* (pp. 1–31). Association for Computing Machinery. [\[CrossRef\]](#)
- Diel, A., Lalg, T., Schröter, I., MacDorman, K., Teufel, M., & Bäuerle, A. (2024). Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers. *Computers in Human Behavior Reports*, 16, 100538. [\[CrossRef\]](#)
- Drolsbach, C., & Pröllochs, N. (2025). Characterizing AI-generated misinformation on social media. *arXiv*. [\[CrossRef\]](#)
- Eppler, M. J. (2006). *Managing information quality: Increasing the value of information in knowledge-intensive products and processes*. Springer. [\[CrossRef\]](#)
- Eppler, M. J., & Wittig, D. (2000). Conceptualizing information quality: A review of information quality frameworks from the last ten years. In *Proceedings of the 2000 conference on information quality (ICIQ 2000)* (pp. 83–96). Massachusetts Institute of Technology (MIT).
- Fischhoff, B. (2013). The sciences of science communication. *Proceedings of the National Academy of Sciences*, 110, 14033–14039. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies* (pp. 167–194). MIT Press. [\[CrossRef\]](#)
- Herrero-Diz, P., & López-Rufino, C. (2021). Libraries fight disinformation: An analysis of online practices to help users’ generations in spotting fake news. *Societies*, 11(4), 133. [\[CrossRef\]](#)
- Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change*. Yale University Press.
- IPCC. (2023). *Sixth assessment report (AR6): Synthesis report*. IPCC. [\[CrossRef\]](#)
- Jamieson, K. H. (2018). *Cyberwar: How Russian hackers and trolls helped elect a president—What we don’t, can’t, and do know*. Oxford University Press.
- Kahan, D. M. (2017). Misconceptions, misinformation, and the logic of identity-protective cognition. *SSRN Electronic Journal*. [\[CrossRef\]](#)
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. [\[CrossRef\]](#)
- Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: A methodology for information quality assessment. *Information & Management*, 40(2), 133–146. [\[CrossRef\]](#)
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the “post-truth” era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353–369. [\[CrossRef\]](#)
- Liu, Z., Zhang, T., Yang, K., Thompson, P., Yu, Z., & Ananiadou, S. (2024). Emotion detection for misinformation: A review. *Information Fusion*, 107, 102300. [\[CrossRef\]](#)
- López, A., Pastor Galindo, J., & Ruipérez-Valiente, J. A. (2024). Frameworks, modeling, and simulations of misinformation: A systematic literature review. *arXiv*. [\[CrossRef\]](#)
- Misirlis, N., & Munawar, H. B. (2023). From deepfake to deep useful: Risks and opportunities through a systematic literature review. *arXiv*. [\[CrossRef\]](#)
- Monteith, S., Glenn, T., Geddes, J. R., Whybrow, P. C., Achtyes, E., & Bauer, M. (2024). Artificial intelligence and increasing misinformation. *British Journal of Psychiatry*, 224(2), 33–35. [\[CrossRef\]](#) [\[PubMed\]](#)
- Muhammed, T. S., & Mathew, S. K. (2022). The disaster of misinformation: A review of research in social media. *International Journal of Data Science and Analytics*, 13, 271–285. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2019). Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favourability. *Political Behaviour*, 42, 939–960. [\[CrossRef\]](#)
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Press.
- Pennycook, G., & Rand, D. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rodríguez Castillo, N. E., Mendoza Carrera, J. E., Andrade-Vásquez, M. M., & Acosta-Barreno, K. (2026). Media literacy education and misinformation in social media among adolescents: A systematic review and meta-analysis. *Journalism and Media*, 7(2), 71. [\[CrossRef\]](#)

- Scheufele, D. A. (2014). Science communication as political communication. *Proceedings of the National Academy of Sciences*, 111, 13585–13592. [CrossRef] [PubMed]
- Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School Misinformation Review*, 4(5), 1–11. [CrossRef]
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, 177, 1333–1352. [CrossRef]
- State of California Department of Justice. (2024). *Attorney General Bonta warns Californians: AI-generated scams are widespread and tricky to spot*. Available online: <https://oag.ca.gov/news/press-releases/attorney-general-bonta-warns-californians-ai-generated-scams-are-widespread-and> (accessed on 17 March 2026).
- Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103–109. [CrossRef]
- Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger, & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). MIT Press.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285. [CrossRef] [PubMed]
- Trninić, D., Kuprešanin Vukelić, A., & Bokan, J. (2022). Perception of “fake news” and potentially manipulative content in digital media—A generational approach. *Societies*, 12(1), 3. [CrossRef]
- Tufekci, Z. (2017). *Twitter and tear gas: The power and fragility of networked protest*. Yale University Press.
- United Nations. (2023). *Our common agenda: Policy brief 8—Information integrity on digital platforms*. United Nations. Available online: <https://disinfopacific.org/infointegrity/> (accessed on 1 July 2026).
- van Dijck, J. (2013). *The culture of connectivity: A critical history of social media*. Oxford University Press. [CrossRef]
- Vishnupriya, G., Jeriel, K. A., RNS, A., Ajay, G., & J, A. G. (2024). Combating fake news in the digital age: A review of AI-based approaches. In *Proceedings of the IEEE international conference on convergence in technology* (pp. 1–6). IEEE. [CrossRef]
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. [CrossRef] [PubMed]
- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33. [CrossRef]
- Wardle, C. (2018). Understanding information disorder. In *Journalism, “fake news” & disinformation* (pp. 43–54). UNESCO.
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe.
- Weikmann, T., & Lecheler, S. (2022). Visual disinformation in a digital age: A literature synthesis and research agenda. *New Media & Society*, 25, 3696–3713. [CrossRef]
- Zárraga-Rodríguez, M., & Álvarez, M. J. (2014). Does the EFQM model identify and reinforce information capability? *Procedia—Social and Behavioural Sciences*, 109, 716–721. [CrossRef]
- Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys*, 53(5), 109. [CrossRef]
- Zrnec, A., Poženel, M., & Lavbič, D. (2022). Users’ ability to perceive misinformation: An information quality assessment approach. *Information Processing & Management*, 59(1), 102739. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.