



Impacts of the popularization of artificial intelligence on the three principles of information integrity

Impactos de la popularización de la inteligencia artificial en los tres principios de la integridad de la información

Carlos Alberto Ávila Araújo

Universidade Federal de Minas Gerais
carlosaraujoufmg@gmail.com

recibido: 15.04.2026 / aceptado: 22.07.2026

Referencia de este artículo

Araújo, Carlos Alberto Ávila (2026). Impacts of the popularization of artificial intelligence on the three principles of information integrity. *adComunica. Revista Científica de Estrategias, Tendencias e Innovación en Comunicación*, nº32. Castellón de la Plana: Departamento de Ciencias de la Comunicación de la Universitat Jaume I, 73-94. DOI: <http://dx.doi.org/10.6035/adcomunica.9551>

KEYWORDS

Information Integrity; Generative Artificial Intelligence; Disinformation; Accuracy; Consistency; Reliability.

PALABRAS CLAVE

Integridad de la información; inteligencia artificial generativa; desinformación; precisión; coherencia; fiabilidad.

ABSTRACT

Introduction: In recent years, to combat attacks against trends such as disinformation, hate speech, and conspiracy theories, the concept of information integrity has been proposed within multilateral international organizations such as the UN and the G20. However, the impacts brought about by the recent popularization of generative artificial intelligence (GAI) tools are not yet present in the documents that present the concept. **Objective:** This study seeks to analyze to what extent the issues and challenges brought about by GAI impact the promotion of information integrity. **Methodology:** The study was conducted using institutional documents produced in recent years as its object. Among the various elements that make up the concept, its three fundamental principles were chosen: accuracy, consistency, and reliability. For the critical analysis of these principles, Jürgen Habermas's theory of communicative action was used. Following an analysis of recent research on GAI, some of the risks and problems brought about by GAI were identified and systematized, and analyzed in light of Habermas's theory, correlated with the three principles of information integrity. **Results:** In recent years, the popularization of the use of generative artificial intelligence (GAI) has brought new challenges to the promotion of information integrity. Specifically, regarding the three principles, AI has brought aspects such as increased disinterest in truth, compromised cognitive authorities, opacity of content production criteria, algorithmic discrimination, complexity of processes to distinguish truth from falsehood, and compromised cognitive authorities. **Conclusion:** GAI has brought several positive aspects, such as time optimization, increased efficiency, and accessibility to large amounts of content. At the same time, it has brought evidence that threatens information integrity and, consequently, science, democracy, environmental protection, and public health. It is believed that new forms of information mediation are possible to ensure healthier, safer, and more inclusive information ecosystems. Thus, the issues raised by generative artificial intelligence need to be considered in documents and actions that promote information integrity.

RESUMEN

Introducción: En los últimos años, para combatir los ataques contra tendencias como la desinformación, el discurso de odio y las teorías de la conspiración, se ha propuesto el concepto de integridad de la información en organizaciones internacionales multilaterales como la ONU y el

G20. Sin embargo, los impactos derivados de la reciente popularización de las herramientas de inteligencia artificial generativa (IAG) aún no se encuentran reflejados en los documentos que presentan el concepto. Objetivo: Este estudio busca analizar en qué medida los problemas y desafíos que plantea la IAG impactan la promoción de la integridad de la información. Metodología: El estudio se realizó utilizando documentos institucionales producidos en los últimos años como objeto de estudio. Entre los diversos elementos que conforman el concepto, se seleccionaron sus tres principios fundamentales: precisión, coherencia y fiabilidad. Para el análisis crítico de estos principios, se utilizó la teoría de la acción comunicativa de Jürgen Habermas. Tras un análisis de investigaciones recientes sobre IAG, se identificaron y sistematizaron algunos de los riesgos y problemas que esta genera, y se analizaron a la luz de la teoría de Habermas, correlacionándolos con los tres principios de la integridad de la información. Resultados: En los últimos años, la popularización del uso de la inteligencia artificial generativa (IAG) ha planteado nuevos desafíos para la promoción de la integridad de la información. Específicamente, en relación con los tres principios, la IA ha generado aspectos como un mayor desinterés por la verdad, la vulneración de las autoridades cognitivas, la opacidad de los criterios de producción de contenido, la discriminación algorítmica, la complejidad de los procesos para distinguir la verdad de la falsedad y la vulneración de las autoridades cognitivas. Conclusión: La IAG ha aportado varios aspectos positivos, como la optimización del tiempo, una mayor eficiencia y el acceso a grandes cantidades de contenido. Al mismo tiempo, ha generado evidencia que amenaza la integridad de la información y, en consecuencia, la ciencia, la democracia, la protección del medio ambiente y la salud pública. Se cree que es posible desarrollar nuevas formas de mediación de la información para garantizar ecosistemas informativos más saludables, seguros e inclusivos. Por lo tanto, es necesario considerar los problemas que plantea la inteligencia artificial generativa en los documentos y acciones que promueven la integridad de la información.

AUTHOR

Carlos Alberto Ávila Araújo [carlosaraujoufmg@gmail.com] is Associate professor at the School of Information Science at UFMG. He holds a PhD in Information Science from UFMG, with postdoctoral research at the University of Porto and the University of Salamanca. He is currently undertaking postdoctoral research at the University of Lisbon. He was president of EDICIC, the Association for Education and Research in Information Science in Ibero-America and the Caribbean.

ACKNOWLEDGEMENT

The author thanks to the Brazilian National Council for Scientific and Technological Development (CNPq) for the research productivity grant and to the Minas Gerais State Research Support Foundation (FAPEMIG) for the postdoctoral fellowship.

1. Introduction

This text presents the results of research that has been ongoing since 2022, initially focused on the study of disinformation phenomena and related issues (such as hate speech, conspiracy theories, and denialism) that have harmful effects on the digital ecosystem, within the framework of a partnership established between several Brazilian universities and the Brazilian Supreme Federal Court, and the Supreme Federal Court, the highest judicial institution in Brazil. In subsequent years, the research became linked to the theme of information integrity, a proposal to combat phenomena associated with disinformation presented by the United Nations (UN) and the G20, the group of the twenty largest economies in the world, and which, in Brazil, was embraced by a pool of institutions, including the National Council for Scientific and Technological Development, the institution that coordinates the scientific research sector in Brazil, and the National Network to Combat Disinformation.

However, in recent years, a new phenomenon has been structurally changing the way people interact with information. This is the intense popularization of the use of generative artificial intelligence (GAI) tools, which has been occurring since 2022. Institutional documents on information integrity practically do not address GAI. But recent studies on GAI have highlighted aspects and phenomena that are directly related to promoting information integrity. It is in relation to this gap that the research reported in this text was developed.

2. Research objectives and hypothesis

As noted above, the research presented in this text falls within the scope of a series of scientific and institutional efforts to understand phenomena such as disinformation, hate speech, and conspiracy theories, as well as post-truth and the infodemic. Furthermore, these efforts focus on finding ways to combat the harmful effects of such phenomena, such as the erosion of trust in science and democracy, climate change denial, denial of public health policies and treatments, and attacks on human rights.

Amid the possibilities for combating these phenomena, such as supporting fact-checking agencies, regulating digital platforms, and promoting media and information literacy, the idea of information integrity has been developed in recent years. Its main objective is to promote a broad explanatory and propositional framework that allows for the inclusion of actions that are often carried out in a punctual and disconnected manner – such as fact-checking, regulation, and literacy actions.

However, there is a gap in contemporary documents regarding information integrity. They start from scientific analyses of disinformation and related phenomena, consider in their analyses the different digital information environments (search engines, social networks, instant messaging applications, big data, big tech business models), but practically do not contemplate information integrity. This is probably due to the fact that the popularization of information integrity tools is very recent. There is, therefore, a knowledge gap regarding information integrity and information integrity.

The research presented in this text seeks precisely to fill this gap. To this end, a survey was conducted, described in the following topic. The hypothesis that guided this work is that information integrity, despite the set of advantages it brings, also structurally threatens information integrity. In this sense, it is essential that the theorizing about information integrity begins to contemplate information integrity when presenting proposals for action and promotion of information integrity.

3. Methodology

This text presents the results of a research project that has been developing since 2022, but which, in its most recent phase, initiated in 2025, focuses on the impacts of the recent popularization of the use of generative artificial intelligence (GAI) in promoting information integrity.

Thus, as a research method, the procedures adopted in the previous stages of the research were maintained. The first of these is the use of the document analysis method, which had as its empirical object the institutional documents produced by multilateral international organizations, such as the UN and the G20. This choice is justified by two facts. The first is that it was these organizations that introduced the discussion about information integrity into the contemporary debate. Although the expression emerged decades ago, its current meaning, linked to disinformation, arose within the scope of such organizations. The second fact is that there are still few scientific works on information integrity, in the sense used by these organizations.

These institutional documents have many elements, aspects, concepts, and problematizations. For the present study, only the three basic principles of information integrity were chosen: accuracy, consistency, and reliability. The reason for this choice is due to the centrality of these principles, which are present in all documents.

For the analysis of this material, Jurgen Habermas's theory of communicative action was chosen. This choice was made after consulting and tracking different theories present in Sociology, Communication, and Information Science, based on consultation

of consolidated manuals and treatises in these areas. The theory of communicative action was considered the most appropriate for two reasons. The first is the fact that, when presenting the concept of communicative action, the author mentions three other types of action (teleological, normative, and dramaturgical) that can be directly related to each of the three principles of information integrity. The second is that, in Habermas's theory, the emphasis is on the action of subjects, and in institutional documents, this emphasis on action is equally evident. The third reason relates to the objective of the theory. Habermas built his theory based on the idea of a "public sphere," that is, an environment in which different subjects could, through the use of arguments and rationality, promote constructive dialogue and reach consensus on the best ways to manage society. In documents on information integrity, similar ideas often appear, suggesting that digital ecosystems should promote dialogue, the active participation of subjects, and the promotion of consensus for the benefit of all.

The analysis is further complemented by the use of authors and theories related to disinformation, with aspects linked to the three principles of information integrity (truth, ethics, authenticity), and with the use of recent theories and reflections on information integrity, especially those related to its ethical aspects (Innerarity, 2025; Zárate Alcarazo, 2026).

4. The theme of information integrity

In recent years, the dynamics of information creation, distribution, and consumption have been greatly impacted by the widespread presence of false (totally or partially) and misleading (distorted, adulterated, or decontextualized) content. Terms such as disinformation and fake news have become frequent to describe this moment (Wardle & Derakhshan, 2017; Mesquita, 2023). Furthermore, long-standing phenomena such as hate speech, denialism, and conspiracy theories have gained new importance in digital environments (Taguieff, 2021; Peker, 2026). The increased importance of this set of phenomena occurred with events in 2016, such as the electoral process in the United States and the Brexit vote in the United Kingdom, from which the expression "post-truth" was consolidated (Allonnes, 2018). In 2020, with the Covid-19 pandemic and the threats that misinformation brought to public health, the term "infodemic" became established (Naeem; Bhatti, 2020). In subsequent years, many studies denounced and mapped such phenomena in various instances and dimensions of human life (Dalkir; Katz, 2020).

Recently, to shift the debate from a negative perspective of identifying problems to a proactive one, the notion of information integrity has emerged. In fact, the expression information integrity arose more than four decades ago but linked to the idea of

information security in corporate environments (Geisler, Prabhaker & Nayar, 2003). However, in recent years, it has gained another significant meaning, directly related to combating disinformation.

The idea of information integrity gained global prominence in June 2023, with the publication of the document *Our Common Agenda: Policy. Brief 8* by the United Nations. Two other important steps took place to consolidate the proposal for information integrity: the signing of the Digital Global Compact and the consolidation of the five global principles of information integrity by the UN in the document *United Nations Global Principles for Information Integrity: Recommendations for Multi-stakeholder Action*, both of which took place in 2024.

Another multilateral international organization involved with information integrity is the G20. In September 2023, this group launched the *Global Declaration on Online Information Integrity*. In April 2024, a G20 side event took place on the theme of Information integrity and trust in the digital environment. The theme was also present at the 19th G20 Summit, which took place in November 2024. In the same year, the G20 published the document *Possible Approaches to Promoting Information Integrity and Trust in the Digital Environment*.

The OECD also published, in December 2024, *OECD Draft Recommendation on Information Integrity*. The recommendation, which was adopted by the OECD Council on a proposal from the Public Governance Committee (PGC), seeks to establish policies for members to address the threats posed by information manipulation and implement measures that promote information integrity in line with the universal human rights of freedom of opinion and expression. The previous year, the OECD promoted the event *Tackling disinformation: Strengthening democracy through information integrity*.

In 2025, the issue of information integrity gained significant prominence in climate negotiations with the launch of the *Declaration on Climate Change Information Integrity* in November during COP30. The document established several international commitments to combat climate misinformation and promote accurate information. The *Global Initiative for Information Integrity* was also launched at this event, a union of forces to strengthen research and measures to combat disinformation campaigns that delay and compromise climate action. Prior to this, in March, *The Climate Information Integrity Summit*, organized by Conscious, had already taken place. Advertising Network (CAN) and *Climate Action Against Disinformation (CAAD)*. Throughout 2025, several events were organized on the topic, such as the *Regional Conference on Information Integrity in West Africa and the Sahel* (September), a *EU DisinfoLab Annual Conference*, in Slovenia (October), a *Copenhagen Conference on Information Integrity* (November), *Sofia Information Integrity Forum* (November), the *Disinfo Pacific*, pro-

moted by the United Nations Information Centre (UNIC) for the Pacific (November) and the Roundtable on Information Integrity in an Evolving AI Landscape, organized by the UN Department of Global Communications, UNESCO and the Delegation of the European Union to the UN (December).

In the various documents produced on information integrity, there are some common points. First, the use of the expression information pollution to emphasize the fact that an information ecosystem becomes corrosive not only through the spread of misinformation, but also through hate speech, conspiracy theories, distorted information, and other phenomena. Second, the understanding that misinformation occurs not only through false content, but also through the falsification of elements of true content, such as authorship, date, dissemination channel, and context. Third, the fact that misinformation does not exist only in the content of messages, in the intentionality of the producers, but also manifests itself in information flows and in the different information regimes that articulate political, economic, technological, social, and other aspects. In the various documents, it is also common to present the three main elements of information integrity: accuracy, consistency, and reliability.

Official documents from the UN and the G20 call for the scientific community to contribute to the consolidation of the concept of information integrity. This work is part of that proposal, forming part of a broader study on the scientific construction of the concept based on its three main elements, using Jürgen Habermas's theory of communicative action. However, in this work, the objective is to discuss a specific aspect of this issue: the recent popularization of generative artificial intelligence tools.

Artificial intelligence has existed for several decades and has been used in various fields and processes. Even in the informational field, AI has been used for many years in search engines, geolocation systems, translation services, and content recommendation models (Russell & Norvig, 2022). But generative artificial intelligence is somewhat more recent, and its popularization and everyday use by a large number of people dates back to 2022 (Córdova, 2025). This popularization has strongly affected the ways in which information is produced, distributed, accessed, and used (Marcus, 2024). It brings, at the same time, risks and possibilities for promoting information integrity. Thus, it is necessary to consider this new dimension in discussions about information integrity.

5. Impacts of AI on information integrity

The discussion undertaken in this text uses Jürgen Habermas's (1981) theory of communicative action as its foundation. The aim of this theory is to analyze the crisis

of modernity based on the observation that economic and bureaucratic rationality is invading and colonizing the sphere of the lifeworld. For Habermas, this form of rationality is inadequate for the lived social world, leading to a loss of freedom and meaning. As a solution, he proposes the idea of communicative action, a form of action oriented towards understanding.

Before presenting the concept of communicative action, Habermas constructs a framework in which he identifies three forms of action: teleological action, normative action, and dramaturgical action. Finally, he presents communicative action. In the analysis carried out in this text, each of the three types of action is used for each of the three elements of information integrity, as a way to verify the impacts of AI. In the end, communicative action is revisited for a global analysis of the impacts of AI on the promotion of information integrity.

5.1. Accuracy

The first element of information integrity is precision. This refers to the objective world, that is, to reality that exists independently of our will. This reality is of the order of contingency. Sometimes, the term used to describe this element is accuracy – which means the idea that there can be greater or lesser precision of a given content in relation to reality. This element evokes a cognitive nature of information: information refers to a world that exists independently of subjects, and there are different ways of making this reference, with greater or lesser precision. It is the dimension of information as knowledge, what Arendt (1958) calls factual reality. Thus, information largely has a testimonial character (Allonnes, 2018), seeking to say something about reality.

Wilson (1983) argues that we know the world in two ways: firsthand, through direct experience; and secondhand, based on the experience of others, through testimonies, that is, through the production of information. Regarding this second form, Wilson introduces the idea of cognitive authority (or epistemic authority) to identify certain actors possessing credibility in relation to specific subjects. This credibility is built within human interactions, resulting in certain individuals acquiring credibility and influence in a particular sphere of human life.

This same idea is proposed by Fallis (2004), who identifies that much of the knowledge we have of the world comes from the experience of others. Thus, he proposes the need for an epistemology of testimony, that is, the fact that the authority of others comes from their experience. Even if certain content comes from a cognitive authority, it may be false, inaccurate, distorted, or fraudulent. The truth power of a testimony is related to the evidence and facts of the world and the ways of obtaining this evidence, which is what guarantees the reliability of that testimony. Cognitive authorities may be linked to institutions. They are not only the sources of information (authorship),

but also the procedures adopted to generate their content, their reputation (past behaviors and history of reliability), the existence of interests or biases, their agreement with other authorities or sources of information, among others.

Some forms of knowledge are every day, intuitive, spontaneous, immediate, and necessary in the face of lived situations. Others are the product of systematic research and study, less immediate, seeking a certain distance from lived experience, committed to the permanent pursuit of objective, in-depth, and systematic knowledge of reality. This second group is scientific knowledge, which has become the cognitive authority of modern societies. This does not mean that other forms of knowledge cannot provide adequate knowledge about reality – but rather that commitment to facts is not their goal. This component of information integrity is directly related to the concepts of disinformation, which see it as actions that deliberately disseminate false information with the intention of deceiving or confusing people to cause misleading perceptions of reality and thus induce decision-making based on error.

In Habermas's theory of communicative action, this dimension corresponds to teleological action. It is a type of action oriented towards a purpose, that is, aimed at achieving a goal. It is also a cognitive action, meaning it presupposes an understanding of an objective reality, external to the subject, which will serve as the basis for achieving the proposed objectives. The knowledge of this objective reality can be more or less rigorous – in any case, it imposes the condition of truth on the subjects' actions.

The intense use of artificial intelligence by people exacerbates the characteristic of digital environments already observed in social networks and search engines: the filter bubble effect, also called echo chambers. In attempting to customize search results for information and recommendation systems, the algorithms of these environments have begun to organize information in such a way that people are kept in bubbles, each bubble fed by its own information flows. Artificial intelligence services provide answers and content adapted to the desires and interests of users and, moreover, are always organized to align with these desires and interests, providing answers that please their users. This weakens the idea of shared truth and, at the same time, strengthens the idea of each person with "their own truth", that is, with what they choose to believe.

The increased use of AI also occurs in a post-truth scenario, that is, one of growing disinterest in and disdain for truth, accompanied by a valorization of emotions and personal beliefs. Han (2022) identifies a kind of loss of the "impulse for truth", of a "will to truth" on the part of people. This is a phenomenon in which it is not lies that take the place of truth, but rather the annulment of the differences between false

and true, an attack on facticity, on the common, shared world to which people could link their actions. AI, interacting with this scenario, deepens this loss of interest in truth, promoting an erosion of the factual. Instead of shared facts, AI offers parallel and distinct worlds to different users, fragmenting perceptions of reality – precisely because it seeks to maximize the time subjects spend using its tools.

What began with search engines—personalizing results and offering distinct content to each individual—is taking on a much larger scale with AI tools. The content produced and used by people is becoming increasingly attuned to their individuality. The impact of this is that people are increasingly losing the ability to see the world as others see it. And the inability to see reality from another's point of view deepens and reinforces echo chambers.

5.2. Consistency

The second element of information integrity is consistency. It refers to the idea that information is not merely about transmitting messages, but a way of acting in the world and, at the same time, constructing that world. Consistency concerns reality not as contingency, as something independent of subjects, but precisely the opposite: the world that one wants to build, the necessary world, considering the demands of the actors and subjects who live in it. Furthermore, the objectives of the subjects are not the same; there are different and often contradictory interests. This is why, in some documents on information integrity, this element is called “balance”: the idea that different interests should be presented and considered in the search for syntheses that are best for everyone.

The central element of this component is its normative character, that is, the construction of norms to guide human actions. This is what Arendt (1958) calls rational truth, that is, the truth constructed by subjects – the world that subjects want, that is, the world that depends on their action and will. This leads to a discussion about the origin of norms and the degree of awareness of subjects about their arbitrary and negotiated nature. It is the dimension of information as culture. It is possible to identify the processes of genesis and awareness of norms in two ways, one individual and the other social.

Weber (1948) presented three types of socially regulated behaviors. The first is usage, a mode of regulating behaviors through the unconscious repetition of certain actions. The second is tradition, which occurs when this repetition becomes part of a long experience, becoming a habit. In this case, the heritage of ancestors becomes a prescription for future action, and there is a relative awareness, on the part of the actors, of these regularities of past behaviors. The third type is convention. In it, subjects become aware that norms can be intentionally produced, that they are the product

of the subjects' actions – and not the product of divine will or some transcendental force. The regulation of behavior starts from an unconscious and unreflective level, followed by a gradual awareness of the norms on the part of the actors.

A second possibility for analyzing norms, from a psychogenetic perspective, is offered by Piaget (1932). Piaget identifies four stages of a child's moral development: pre-moral, heteronomous morality, intermediate, and moral autonomy. In the first, the child does not know the social rules and acts in the world in a relatively unconscious way, without understanding or accepting the rules. In the second, the child understands the norms as existing in themselves, external to their consciousness and will. In the third, the child ceases to rigidly obey the norm, observing it in specific situations and interpreting it. In the final stage, the child recognizes the consensual nature of the norm, perceives that its necessity stems from the relationship with the other, and sees that its validity is conditioned by the reciprocal respect of the parties – the norm can be renegotiated.

Thus, norms arise both from a social dynamic imposed upon individuals and from individuals who actively construct them. Furthermore, norms are not given, they are not immutable; they are dynamic and negotiable between the parties. They are also spaces of autonomy for individuals, of democracy, and, at the same time, political, since they presuppose the need to regulate the coexistence of people. This is the point where Habermas works with his proposal for communicative action. In his discussion of moral consciousness and communicative action, he also draws on Weber and Piaget, but his starting point is different: Kohlberg's (1981) theory of the development of moral consciousness, which identifies the existence of three levels and six stages. The first level is the pre-conventional, which includes the stage of punishment and obedience (literal obedience to authority to avoid punishment) and the stage of individual instrumental objective and exchange (following the rules to meet one's own interests).

The second level is the conventional one, which includes the third stage of mutual interpersonal expectations (when it is recognized that others also have interests), and the fourth, preservation of the social system and conscience (when one seeks to fulfill one's duty in society to maintain its well-being). The third level is the post-conventional or principle-based one. In this level, the decisions of individuals are based on values and principles with which all individuals in a society agree. It includes the fifth stage (of original rights and the social contract, that is, when one seeks to uphold the rights, values, and contracts of a society) and the sixth, of universal ethical principles, which presupposes the existence of universal principles that all humanity should follow. It is important to emphasize that Kohlberg understands the passage from one stage to another as a learning process. Habermas relates Kohlberg's six

stages of moral judgment to the discourses and communicative acts of individuals to support his theory of communicative action.

Thus, consistency relates to the intersubjective world, that is, to the link between an action and the objectives it proposes. This element presupposes the idea that information is not merely something linked to the emission of messages but is a way of acting in the world and, at the same time, constructing that world. Consistency concerns reality not as contingency, as something independent of subjects, but precisely the opposite: the world that one wants to build, the necessary world, considering the demands of the actors and subjects who live in it. Landmarks of this element are the Universal Declaration of Human Rights of 1948 and, more recently, the Sustainable Development Goals (SDGs), a collection of 17 global goals established by the UN General Assembly involving issues such as poverty, hunger, health, education, global warming, and gender equality.

In Habermas's theory, this dimension refers to normative action. It relates to a type of action oriented towards certain common values, shared among a group of people in which the subject of the action lives. In a social group, it is expected that the norms of coexistence will be obeyed by all its members. This action has an intersubjective nature, related to the moral correctness of the subjects, that is, to their conformity to the norms.

Regarding this aspect, it is important to highlight, first of all, the opacity of AI systems, that is, the fact that they operate with criteria that are often unknown to produce content. There is little or no transparency about the data set used in their training, the contexts applied, or the reliability of the processed information. Because of this lack of traceability and epistemological control, it is common for such models to produce incorrect or decontextualized responses, hallucinations. Transparency, therefore, has an ethical value, insofar as making decisions and processes visible to the public guarantees their scrutiny, accountability, the fight against biases, and the possibility of improving services (Corrado, 2026).

However, beyond opacity, it is known that AI technologies operate with certain elements, such as the use of more significant information present in the digital environment. This is a complicated element, insofar as popular content, with high visibility and engagement, does not tend to be the highest quality or most reliable. On the contrary, there is a lot of content based on hatred, prejudice, discrimination, and violence, which can be used to fuel the production of AI-generated content (Han, 2022). In conjunction with the previous factor, hatred acts to ensure that the other is seen as an enemy, as someone to be destroyed – which precisely prevents dialogue and the possibility of consensus building.

In other words, there are significant risks of biases that contradict multilaterally established principles. Many studies show that search engines, social media recommendation mechanisms, and now AI (Artificial Intelligence) tools employ racist, misogynistic criteria, or criteria that encourage violence, war, or scientific or historical denialism (O'Neil, 2018; Empoli, 2019). Another major concern regarding the use of Artificial Intelligence is that it may perpetuate, and perhaps even increase, the phenomenon of epistemic injustice (Fricker, 2017), that is, the unequal opportunity for different subjects and cultures to produce, transmit, and disseminate knowledge. The algorithmic opacity typical of AI tools may end up privileging only those sources of information that are already hegemonic in the geopolitical and commercial sphere, since Artificial Intelligence mechanisms draw on data collected from multiple sources and may reproduce existing asymmetries in the digital informational environment. This concern was translated and detailed in UNESCO's 2022 publication, *Recommendation on the Ethics of Artificial Intelligence*.

AI biases tend to be related to the data used for training, but also to the intentions and limitations of its programmers and developers. In this sense, many studies have shown that AI tools are produced by technology companies (the big techs) that are capitalist companies, committed to profit, and not necessarily concerned with ethics, democracy, justice, and inclusion (Zuboff, 2019; Varoufakis, 2023). This is aggravated by the fact that there are few companies based practically in a single country, and with a rejection of the regulatory frameworks of the other countries in which they operate (Silveira, Souza & Cassino, 2021; Faustino & Lippold, 2023; Valente & Grohmann, 2024). In other words, AI is not neutral – and a challenge is precisely to have an AI that follows principles of honesty, respect for all, and the pursuit of collective well-being, and not just that of certain groups or individuals.

5.3. Reliability

Finally, the third element of information integrity is reliability. It concerns the subjective dimension of information, that is, the producer of a given message, the completeness of its content, and the elements that compose it: the channel of its dissemination, the date, and the context in which it was produced. Unlike the first two elements, which relate to the objective world of facts or the intersubjective world of values, here the reference is to the document itself, the informational content itself – it refers to the truth of the document. This means ensuring that what is being presented as part of an information action (with a specific author, date, and content) is, in fact, what it claims to be – that is, it is not pretending to be something else. This is why, in some documents on information integrity, the expression used is “authenticity”, that is, the fact that each of the elements that make up the information is true in relation to the information itself. This refers to the idea of truth, not the truth of the world, but the truth of the document, the record of knowledge, as it was produced and disseminated. It

is thus opposed to the idea of falsification, adulteration, or any fraudulent alteration. In some documents, the term used to designate this attribute is “completeness”, in the same sense, that is, the preservation of the elements that compose a given record. It refers to that which is whole, which has not suffered alteration, breakage, or distortion; that is, it is preserved as produced by its author and disseminated through a given channel. It is the idea of information as communication.

Thompson (1990) uses this expression to designate a wide range of specifically human products (actions, objects, expressions) that are meaningful, that is, that involves a creative action related to symbolic capacity. He presents what he calls five aspects that compose these symbolic forms, that is, that are involved in their constitution, in their very form of existence. The first is the intentional aspect, that is, the fact that these symbolic forms are expressions of a subject. The second is the conventional aspect, which refers to the fact that both the production and interpretation of symbolic forms involve the application of certain rules, codes, and conventions. The third is the structural aspect, related to the fact that symbolic forms possess a certain articulated structure, are composed of elements that relate to each other – what is commonly called content. The fourth aspect is the referential aspect, that is, the fact that symbolic forms represent something, refer to something, point to something outside of themselves, which can be objects, people, or situations. The fifth aspect is contextual, which relates to the fact that symbolic forms are always embedded in specific socio-historical contexts, and it is within these contexts that they are produced, circulate, and are interpreted.

In his discussion, Thompson explains that the meaning of symbolic forms is not contained in one aspect or another. It is not equivalent to the producer’s intention, the convention in which it fits, the internal structure of the messages, the objects to which it refers, or its context. Meaning is complex; it is a product of the interaction of all these elements, which act together for the completeness of an act of communication. And it is for this reason that the alteration or falsification of any of them compromises its completeness and, consequently, its integrity.

In Habermas’s theory, this element corresponds to dramaturgical action, in which the action is a form of representation (or self-representation) of the subjects, that is, an expression before an audience. In this sense, the action has a subjective reference, that is, it is an act that involves the subject producing the act and the circumstances in which it happens – the context. This type of action thus presupposes sincerity and authenticity as a constituent element.

Thus, regarding reliability, the first aspect to consider is that the authorship of messages is frequently lost in the production of content automatically generated by AI.

Along with authorship, institutional backing can also be lost, making it difficult to verify the accuracy of the content, that is, to assess its relationship with the reality of the facts. The very notion of cognitive authority is at risk.

Reliability also refers to that which is whole, which has not suffered alteration, breakage, or distortion; that is, which is preserved exactly as produced by its author and disseminated through a specific channel. And here lie other risks associated with the widespread dissemination of Generative Artificial Intelligence. Because, in addition to losing the correct authorship of content, there can be a mixing of content from different sources, losing the completeness that associates content with authors, channels, and especially contexts. Much information produced by Artificial Intelligence is disseminated in a decontextualized way or disconnected from its production conditions, which are fundamental to determining its possible meanings.

There is another element. Images, audio, and video have always been evidence of a fact, indicative content that such a fact occurred – precisely because they are a product of that fact. With AI, images, audio, and videos are produced that do not correspond to reality. Although the possibility of this occurring has always existed, AI increases sophistication, scale, and ease, as well as reducing the costs and barriers to doing so (Herzog, 2024). This implies the need for people to be willing to check and verify if something is real when they receive audio, video, or images. At the same time, it is unfeasible to imagine a world in which it would be necessary to check everything all the time. It is the very notion of trust that changes in the algorithmic scenario of AI, in a world where falsification is confused with reality. If the notion of infodemic highlights the difficulty people have in identifying true and false information in a scenario of information abundance, with AI this scenario is aggravated, given inauthentic content that simulates reality. What people see, hear, or read is no longer a guarantee that something is real or has happened. Botsman (2017) points to this shift in trust, from a local (community) dimension to an institutional one and now to a distributed trust, with a strong algorithmic bias, in which technology changes our own attitude towards trust. With deepfakes, the situation worsens, with manipulated content becoming increasingly indistinguishable from reality. Thus, in a climate of suspicion where everything can be manipulated, an erosion of trust occurs.

But engagement on social media and interactions themselves can also become simulations, as AI can generate bots to like content, increase its visibility, and even interact in the comments – for example, giving a false dimension and importance to a problem, an issue, or a political group.

6. Discussion

The analysis undertaken resulted in the identification of the impacts of algorithmic information on each of the three basic principles of information integrity. Thus, accuracy, which can be understood from Habermas’s concept of teleological action, is affected by the increased disinterest in factual truth, the growth of echo chambers, and the isolation of subjects in their own “truths.” Consistency, analyzed from Habermas’s concept of normative action, is affected by algorithmic opacity, polarization, discriminatory biases, and epistemic injustice. Finally, reliability, related in the text to Habermas’s dramaturgical action, is affected by the loss of authorship and its respective institutional backing, decontextualization, and deepfakes.

The results can be seen below in Figure 1:

Figure 1.

Type of action, according to Habermas’ theory	Principle of information integrity related	Impacts of GAI
Teleological	Accuracy	• disinterest in the truth • strengthening of echo chambers • isolation of subjects
Normative	Consistency	• opacity of platforms • polarization • biases of big tech • epistemic injustice
Dramaturgical	Reliability	• loss of authorship and institutional backing • loss of context • deep fake

Source: compiled by the author.

Returning to Habermas, it is important to highlight that, after presenting the three types of action, the author develops the concept of communicative action, which is geared towards understanding between subjects capable of speaking and acting. In a way, it integrates the three previous types, since communicative action refers to the objective world, the intersubjective world, and the subjective world. However, in communicative interaction, the validity of each of these references can be contested by the interlocutors. For this reason, Habermas presents what he calls four validity claims. The first of these, intelligibility, is a presupposition of the others. The other three are directly related to the three types of action presented by the author. Teleological action has a claim to truth, that is, a correctness or adequacy to what

the world is. Normative action has a claim to correctness, that is, to conformity to the norms of a group. Finally, dramatic or expressive action has a claim to truthfulness, that is, it seeks to be authentic, a true representative of the identity and aspirations of the subjects.

As pointed out in previous sections, Artificial Intelligence is building new ways of producing, circulating, using, and appropriating information in contemporary societies. This also means new forms of validity claims, that is, new criteria for truth, ethics, and authenticity. This poses new challenges related to the veracity, consistency, and reliability of information, in addition to those already foreseen in documents on information integrity.

Regarding issues related to information integrity, recent discussions stand out concerning the business model of AI production, the planetary scale of its operation, the need for regulation, and the necessary stances, on the part of individuals, for a responsible and critical use of Artificial Intelligence tools, considering Artificial Intelligence as a new form of agency that can be leveraged for the benefit of humanity and the environment (Unesco, 2022; Floridi, 2023).

Information has the potential to build negotiations and consensus, to produce more rigorous knowledge, and to construct civilizations that promote justice. This potential is like Habermas's concept of a public sphere, that is, a space where different social actors, using reason and discourse, present their ideas and deliberate together on what is best for everyone. This concept was developed in 1962. Sixty years later, the author updated it for the current scenario in which the public sphere largely takes place in the digital environment (Habermas, 2022). Now, in addition to the challenges of the digital environment, there are those brought about by GAI.

7. Conclusion

Many recent studies on AI have highlighted its potential for optimizing tasks and processes, saving resources and time, increasing the quality of products and services, and even combating misinformation through the detection of false content and identification of patterns of discrimination (AbuJarour, Qarariah, Saadeh & Salem, 2024; Saeidnia et al., 2025). However, as discussed in the previous section, AI also presents challenges for promoting information integrity, given its business model and production infrastructure, as well as its interaction with important phenomena such as the recent devaluation of truth (the so-called post-truth) and the infodemic (the difficulty people have in distinguishing true from false information).

One of the concepts that has been used to analyze contemporary informational dynamics is that of disintermediation. Lind (2003) argues that the term refers to the phenomenon through which content can be disseminated and accessed by people without traditional intermediaries such as mass media. Giacomini (2023), however, argues that the internet has not effectively brought about disintermediation. It only replaces the dominant mediation of the 20th century, that of mass media, with another. Thus, the author proposes the concept of “neo-intermediation”.

In Araújo (2023), there is an analysis of the concept of information disintermediation from a critical perspective. The author points out the idea of disintermediation as an illusion that some people have that certain content reaches them in a disintermediated way (for example, content that arrives in instant messaging application groups or through search engines) while others would be accessed through mediation. This type of view ignores the fact that search engines have criteria for prioritizing results, that social networks have recommendation systems, and that instant messaging application groups are organized by political, economic, or religious groups and actors. In this sense, IAG deepens this illusion of disintermediation, since users come into contact with content produced without marks of authorship and context.

Combining this reflection with the findings of the research reported in this text, it is noted that it is essential to rethink and mature the criteria through which knowledge is produced and used by people. Theories of knowledge (Nagel, 2014) emphasize the role of the first-person point of view and also of testimonies. Thus, instead of no longer being necessary, informational mediations become even more important in a scenario of intense use of information technology – whether they are journalistic mediations (Matos Mejías & Carrasco Polaino, 2025), scientific mediations (Miranda et al, 1999), library, archival and museological mediations (Araújo, 2023) or others. Thus, a future research agenda needs to contemplate the new challenges for traditional knowledge mediators.

At the same time, such challenges renew the very meaning of ethics in the field of information, challenging the traditional theories and assumptions of both virtue ethics and deontology and utilitarianism, with the question of truth gaining prominence (Blackburn, 2021). Another field that deserves highlighting is that of literacy in AI (Zimmerman, 2018; Yi, 2021; Ng et al, 2021; Williamson et al, 2023; Thirunavukarasu et al, 2023). Initiatives have been developed, both in terms of empowering people and promoting a critical stance, as well as democratization and responsible use.

Artificial intelligence is intertwined with the political, social, cultural, economic, legal, and technological dimensions of informational dynamics (Suleyman & Bhaskar, 2023). Therefore, artificial intelligence is a topic that needs to be considered in studies on the

conceptualization and promotion of information integrity. This text aims to contribute to this understanding, starting from the seriousness of the phenomena of disinformation that have strongly impacted democracy, public health, environmental protection, human rights, and other dimensions of human life across the planet.

References

- AbuJarour, S., Qarariah, A., Saadeh, N., Salem, M. (2024). AI, misinformation, and fake news: a literature review of ethical and technical approaches. In: Mansour, N. & Bujosa Vadell, L.M. (Eds). *Finance and law in the metaverse world: contributions to finance and accounting*. Cham: Springer.
- Araújo, C. A. A. (2024). Disintermediation of information: A new phenomenon of study for information science. In: *Annals of Library and Information Studies*, v. 70, n. 4, 189-196.
- Allonnes, M. (2018). *La faiblesse du vrai : ce que la post-vérité fait à notre monde commun*. Paris: Seuil.
- Arendt, H., 1958. *The human condition*. Chicago: The University of Chicago Press.
- Blackburn, S. (2021). *Ethics: a very short introduction*. Oxford: Oxford University Press.
- Botsman, R. (2017). *Who can you trust? How technology brought us together*. London: Penguin.
- Córdova, P. (2025). *Inteligência artificial: entre o fascínio e o medo*. São Paulo: Contexto.
- Coroado, S. (2026). *Transparência na vida pública*. Lisboa: Fundação Francisco Manuel dos Santos.
- Dalkir, K. & Katz R. (Eds.). (2020). *Navigating fake news, alternative facts, and misinformation in a post-truth world*. Hershey: IGI Global.
- Empoli, G. (2019). *Les ingénieurs du chaos*. Paris: JC Lattès.
- Fallis, D. (2004). On verifying the accuracy of information: philosophical perspectives. In: *Library Trends*, v.52, n.3, 463-487.
- Faustino, D. & Lippold, W. (2022). *Colonialismo Digital: por uma crítica hacker-fanoniana*. São Paulo: Raízes da América.
- Floridi, L. (2023). *The ethics of artificial intelligence: principles, challenges, and opportunities*. New York: Oxford University Press.
- Fricker, M. (2007). *Epistemic injustice: power and the ethics of knowing*. New York: Oxford University Press.
- Geisler, E, Prabhaker, P. & Nayar, M. (2003). Information integrity: an emerging field and the state of knowledge. In: *PICMET '03: Portland International Conference on Management of Engineering and Technology Technology Management for Reshaping the World*. Portland, 217-22.
- Giacomini, G. (2023). Disintermediation and/or Neointermediation? The "Fourth Power" of Small and Large Intermediaries in Digital Public Sphere. In: *Polish Sociological Review*, v. 222, 249-262.
- Habermas, J. (1981). *The theory of communicative action vol. 1: reason and the rationalization of society*. Boston: Beacon Press.

- Habermas, J. (2022). *Ein neuer Strukturwandel der Öffentlichkeit und die deliberative Politik*. Berlin: Suhrkamp Verlag.
- Han, B.-C. (2022). *Infocracy: digitization and the crisis of democracy*. Cambridge: Polity Press.
- Herzog, W. (2024). *Die zukunft der wahrheit*. Munich: Verlag München.
- Kohlberg, L. (1981). *Essays on moral development, Vol. I: The philosophy of moral development*. New York: Harper & Row.
- Innerarity, D. (2025). *Una teoría crítica de la inteligencia artificial*. Barcelona: Galaxia Gutenberg.
- Lind, R. (2003). Disintermediation. In: *Encyclopedia of new media* (p. 151-151). SAGE Publications, Inc. DOI: <https://doi.org/10.4135/9781412950657.n76>.
- Matos Mejías, C., & Carrasco Polaino, R. (2025). Implementación de la Inteligencia Artificial en los estudios de Periodismo de la Facultad de Ciencias de la Información de la Universidad Complutense de Madrid. In: *European Public & Social Innovation Review*, v. 10, 1-18. DOI: <https://doi.org/10.31637/epsir-2025-1136>.
- Marcus, G. (2024). *Taming Silicon Valley: how we can ensure that AI works for us*. Cambridge: MIT Press.
- Miranda, G.F. et al (1999). Scientific Information Strategies: "Disintermediation" or a New Way of Intermediation?. In: Bakker, S. (Ed.) *Libraries without Limits: Changing Needs — Changing Roles*. Springer, Dordrecht.
- Nagel, J. (2014). *Knowledge: a very short introduction*. Oxford: Oxford University Press.
- Naeem, S. B. & Bhatti, R. (2020). The Covid-19 'infodemic': a new front for information professionals. In: *Health Information and Libraries Journal*, v. 37, n. 3, 233-239.
- Ng, D. T. K., et al (2021). Conceptualizing AI literacy: An exploratory review. In: *Computers and Education: Artificial Intelligence*, n. 2, 100041. DOI: <https://doi.org/10.1016/j.caeai.2021.100041>.
- Peker, L. (2026). *La odioocracia*. Madrid: Libros del KO.
- O'Neil, C. (2018). *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown.
- Piaget, J. (1932). *Le jugement moral chez l'enfant*. Paris: Presses Universitaires de France.
- Russell, S. & Norvig, P. (2022). *Artificial intelligence: a modern approach*. Harlow: Pearson.
- Saeidnia, H. et al. (2025). Artificial intelligence in the battle against disinformation and misinformation: a systematic review of challenges and approaches. In: *Knowledge and Information Systems*, v. 67, 3139-3158.
- Silveira, S., Souza, J. & Cassino, J. (Orgs). (2021). *Colonialismo de dados*. São Paulo: Autonomia literária.
- Suleyman, M. & Bhaskar, M. (2023). *The coming wave: technology, power, and the twenty-first century's greatest dilemma*. New York: Crown.
- Taguieff, P.-A. (2021). *Les théories du complot*. Paris: Humensis.
- Thompson, J. B. (1990). *Ideology and modern culture: critical social theory in the era of mass communication*. Redwood: Stanford University Press.
- Thirunavukarasu, A. J. et al (2023). Democratizing artificial intelligence imaging analysis with automated machine learning: Tutorial. In: *Journal of Medical Internet Research*, v. 25. DOI: <https://doi.org/10.2196/49949>.

- Unesco - United Nations Educational, Scientific and Cultural Organization (2022). *Recommendation on the ethics of artificial intelligence*. Paris: Unesco.
- Valente, J. & Grohmann, R. (2024). Critical data studies with Latin America: Theorizing beyond data colonialism. In: *Big Data & Society*, v. 11, n. 1.
- Varoufakis, Y. (2023). *Technofeudalism: what killed capitalism*. Vintage Publishing.
- Wardle, C. & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Strasbourg: Council of Europe.
- Weber, M. (1948). *Essays in sociology*. London: Taylor & Francis.
- Williamson, B., et al (2023). Critical perspectives on AI in education: Political economy, discrimination, commercialization, governance and ethics. In: B. Du Boulay, A. Mitrovic, & K. Yacef (Eds.), *Handbook of Artificial Intelligence in Education* (pp. 553–570). Cheltenham: Edward Elgar Publishing.
- Wilson, P. (1983). *Second-hand knowledge: an inquiry into cognitive authority*. Westport: Greenwood.
- Zárate Alcarazo, L. (2026). *Ética en la inteligencia artificial*. Molins de Rei: Filosofia&Co.
- Zimmerman, M. R. (2018). Teaching AI: Exploring new frontiers for learning. International Society for Technology in Education. Portland: Ebsco.
- Zuboff, S. (2019). *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. New York: PublicAffairs.