Journal on Interactive Systems, 2026, 17:1
ISSN: 2763-7719 • doi: 10.5753/jis.2026.7388

Research Paper

Evolving Ethical Communicability in Generative AI Interfaces: A Longitudinal Comparative Analysis of Ethical Principles Across System Versions

Libiane Gomes [Federal University of Viçosa | *libiane.gomes@ufv.br*]
João Carlos Santana Silveira [Federal University of Minas Gerais | *joaocss@ufmg.br*]
Helena Martins [Federal University of Viçosa | *helena.martins@ufv.br*]
Cristiane Aparecida Lana [Federal University of Lavras | *cristiane.lana@ufla.br*]
Maria Lúcia Bento Villela [Federal University of Viçosa | *maria.villela@ufv.br*]

Departamento de Informática, Av. Peter Henry Rolfs, s/n, Campus Universitário, Viçosa, MG, 36570-900, Brazil.

Abstract. *Background:* Generative Artificial Intelligence (AI) systems have become central interactive artifacts in activities such as information seeking, creative work, education, and decision support. As these systems increasingly mediate human-computer interaction, concerns regarding transparency, accountability, autonomy, fairness, and social impact have intensified. Although ethical principles for AI are well established in regulatory documents and conceptual frameworks, prior research has shown that their communication through interface elements is often inconsistent and opaque. Furthermore, the rapid evolution of generative AI systems raises challenges for understanding ethical communicability as an evolving, rather than static, property of interactive systems. *Purpose:* This paper investigates how ethical communicability in generative AI systems evolves over time by conducting a longitudinal comparative analysis of successive system versions. Extending prior work presented at IHC 2025, the study compares recent and earlier versions of three widely used generative AI systems, examining changes in how ethical principles are communicated at the interface level. The goal is to identify patterns of continuity, improvement, and regression, contributing empirical and conceptual insights to the design and evaluation of generative AI–based interactive systems. *Methods:* The study adopts the Semiotic Inspection Method (SIM) in a scientific context, supported by a semiotics-based epistemic tool for ethical reflection, to analyze how ethical principles are communicated through interface signs. The analysis is guided by the AI4People ethical principles of Beneficence, Non-Maleficence, Autonomy, Justice, and Explicability. Using the same inspection protocol, ethical scenarios, and analytical categories employed in the original study, we re-inspected the current versions of the three systems. The results from both inspection phases were systematically compared using a structured analytical framework, enabling the identification of additions, refinements, and inconsistencies in ethical communication across versions. Researcher triangulation was applied to ensure analytical rigor and consistency. *Results:* The longitudinal comparison reveals that ethical communicability in generative AI systems is dynamic and uneven. Some systems show incremental improvements, particularly in interface elements related to explicability and risk mitigation, such as clearer disclaimers, refined feedback mechanisms, and expanded data control options. However, other ethical principles - especially beneficence and justice - remain weakly communicated or largely implicit across versions. The analysis also exposes persistent inconsistencies between metalinguistic commitments expressed in policies and the ethical cues available during interaction. These findings suggest that system updates do not necessarily lead to systematic or holistic improvements in ethical communication, but rather to fragmented and principle-specific changes. *Conclusion:* By shifting from a static to a longitudinal perspective, this study demonstrates the value of analyzing ethical communicability as an evolving property of generative AI interfaces. The findings highlight that updates to generative AI systems can both strengthen and undermine the communication of ethical principles, underscoring the need for continuous and systematic ethical evaluation in the design of interactive systems. The paper contributes empirical evidence on how ethical communication changes over time, methodological insights into the use of SIM for longitudinal analysis, and design implications for fostering more transparent, accountable, and human-centered generative AI systems. This extended investigation reinforces the relevance of ethical communicability as a core concern for HCI research and practice.

Keywords: Ethical Communicability, Semiotic Inspection Method, Generative AI Systems, Longitudinal Analysis

Edited by: Milene Silveira | **Received:** 13 January 2026 • **Accepted:** 19 May 2026 • **Published:** 23 July 2026

1 Introduction

Generative Artificial Intelligence (AI) systems have rapidly become central interactive artifacts in contemporary digital ecosystems, supporting activities such as information seeking, creative production, education, and decision support. Through conversational interaction, these systems enable users to generate text, images, code, and other content, reshaping how people engage with technology and delegate cognitive and creative tasks. As their adoption grows, however, generative AI systems intensify ethical concerns related to transparency, accountability, autonomy, fairness, and social impact. Their outputs may reproduce social biases, generate misleading or harmful information, violate privacy expectations, or obscure responsibility for automated decisions, particularly in sensitive contexts [Nissenbaum, 1996; Johnson, 2004; Shneiderman, 2020; Bender *et al.*, 2021; Brey and Dainow, 2024].

In response to these challenges, a growing body of work

has proposed ethical principles, guidelines, and regulatory frameworks to guide the responsible development of AI systems. Initiatives such as the AI4People framework [Floridi *et al*., 2018] articulate core ethical principles - Beneficence, Non-Maleficence, Autonomy, Justice, and Explicability - that have become central references in discussions on responsible AI. While these principles play an essential normative role, they are often articulated at a high level of abstraction, offering limited guidance on how ethical commitments should become visible and meaningful to users during interaction [Shneiderman, 2020; Morley *et al*., 2021; Johnson and Smith, 2021].

From a Human-Computer Interaction (HCI) perspective, this limitation points to a critical gap between ethical intentions and user experience. Ethical considerations are frequently communicated through policies, terms of service, or external documentation, rather than being embedded in interface elements, interaction flows, and system behaviors. As a result, users may encounter ethical commitments in a fragmented or indirect manner, requiring them to infer system limitations, risks, and responsibilities on their own, thereby undermining informed use and trust [Binns *et al*., 2018; Shneiderman, 2020]. This gap raises concerns not only about whether ethical principles are adopted, but also about how effectively they are communicated and sustained through interaction.

In interactive systems, ethics may be framed as a matter of communicability, that is, as something that is conveyed and interpreted through interaction rather than merely defined by abstract principles or external policies. From this perspective, ethical commitments are expressed through interface elements, interaction flows, and system behaviors, shaping how users understand risks, responsibilities, and limitations during use [De Souza, 2005]. Studies in Value Sensitive Design further emphasize that ethical values are not only embedded at design time but are enacted and interpreted in situated contexts of use, often in uneven and context-dependent ways [Friedman and Hendry, 2019]. In AI-based systems, prior work has shown that principles such as transparency, fairness, and accountability are frequently communicated implicitly and inconsistently, requiring users to infer ethical implications from fragmented cues [Binns *et al*., 2018; Morley *et al*., 2021]. Together, these perspectives suggest that ethical communication is not an all-or-nothing condition, but rather a quality that can vary in degree, coherence, and integration within interactive systems.

Our previous study, presented at IHC 2025, investigated how generative AI systems communicate ethical considerations to users through interface elements, adopting a semiotics-based perspective grounded in the Semiotic Inspection Method (SIM) [Gomes *et al*., 2025]. By analyzing three widely used generative AI systems, that study revealed inconsistencies and opacity in how ethical principles were communicated to users, highlighting tensions between stated ethical commitments and the interactional cues available during use. While the study demonstrated the relevance of SIM for examining ethical communicability in generative AI interfaces, it also focused on a specific set of system versions, offering a snapshot of ethical communication at a particular moment in time.

Generative AI systems, however, are not static artifacts. They evolve continuously through interface redesigns, policy revisions, and changes to underlying models. These changes may strengthen, weaken, or reconfigure how ethical principles are communicated to users. Understanding ethical communicability as a dynamic property of interactive systems, one that may progress unevenly, stagnate, or regress over time, thus becomes a key challenge for HCI research.

To address this challenge, this paper extends our previous work [Gomes *et al*., 2025] by adopting a longitudinal and comparative perspective on ethical communicability in generative AI interfaces. We analyze the current versions of the same three systems previously studied and systematically compare them with earlier versions, applying the same analytical framework, ethical scenarios, and inspection protocol. By examining additions and refinements in ethical communication across system versions, we seek to identify patterns that indicate different degrees of integration, consistency, and operationalization of ethical principles over time.

Methodologically, the study employs the SIM in a scientific context [de Souza and Leitão, 2009], supported by a semiotics-based epistemic tool for ethical reflection [Barbosa *et al*., 2021]. The analysis is guided by the AI4People ethical principles: Beneficence, Non-Maleficence, Autonomy, Justice, and Explicability [Floridi *et al*., 2018], allowing for a structured comparison of ethical communication across time. This longitudinal application of SIM enables us to examine ethical communicability not merely as a design snapshot, but as an evolving characteristic shaped by ongoing design and development decisions.

The contributions of this paper are threefold. First, it provides new empirical evidence on how ethical communication in generative AI systems changes over time, revealing uneven and fragmented trajectories across principles and systems. Second, it advances methodological discussions by demonstrating the applicability and limitations of SIM for longitudinal analyzes of non-deterministic conversational AI systems. Third, it offers design-oriented insights and implications that support more systematic, transparent, and human-centered communication of ethical considerations in generative AI interfaces.

This research also contributes to broader discussions on Ethics and Responsibility [Rodrigues *et al*., 2024], as well as the Implications of Artificial Intelligence in HCI [Duarte *et al*., 2024], as articulated in the Grand Research Challenges in HCI in Brazil for 2025-2035 (GranDIHC-BR) [Pereira *et al*., 2024]. By foregrounding ethical communicability as a dynamic property of interactive systems, the study reinforces the need for continuous ethical evaluation and design practices that evolve alongside generative AI technologies.

The remainder of this paper is organized as follows. Section 2 presents the theoretical background, including the SIM and ethical approaches to AI design. Section 3 describes the longitudinal methodology and comparative analysis procedure. Section 4 presents the results of the comparative inspection across system versions. Section 5 discusses the findings and their implications for design and evaluation. Finally, Section 6 concludes the paper and outlines directions for future research.

2 Background and Related Work

This section presents the theoretical and methodological foundations for analyzing ethical communicability in generative AI interfaces. We review principle-based approaches to responsible AI and discuss the gap between abstract ethical commitments and their communication at the interface level before introducing the SIM as a lens for analyzing interaction as designer-user communication. Building on this foundation, we frame ethical communication as an interaction-centered phenomenon, supported by a semiotics-based epistemic tool to examine how ethical responsibilities are conveyed in non-deterministic AI systems.

2.1 Ethical Principles and Responsible AI

The increasing adoption of Artificial Intelligence systems in socially relevant contexts has motivated extensive discussions on ethical principles and responsible AI. Over the last decade, multiple initiatives have proposed normative frameworks to guide the development and deployment of AI systems, aiming to mitigate risks related to discrimination, loss of autonomy, lack of accountability, and social harm [Floridi *et al*., 2018; Shneiderman, 2020; Jobin *et al*., 2019]. These discussions reflect growing concerns about the societal consequences of AI systems that increasingly mediate decision-making, information access, and human agency. Among these initiatives, the AI4People framework stands out for articulating a set of core ethical principles: Beneficence, Non-Maleficence, Autonomy, Justice, and Explicability, which synthesize philosophical, legal, and socio-technical concerns into a coherent ethical reference for AI systems [Floridi *et al*., 2018]. Such principles are described below:

- **Beneficence** - it consists of creating systems that benefit humanity, promoting well-being for all people directly affected by them in the same way and, consequently, for the planet;
- **Non-Maleficence** - it consists of creating systems that do not harm people and avoiding the damage that may arise from the excessive use and misuse of AI systems, such as preventing violations of people's privacy and security;
- **Autonomy** - it consists of continually granting human beings the power to decide which decisions they should make themselves and which should be made by AI systems;
- **Justice** - it consists of using AI systems to correct past mistakes, such as eliminating unfair discrimination, creating benefits that can be shared by society, and preventing the creation of new harms;
- **Explicability** - it consists of providing AI systems with intelligibility (answering the question "How does the system work"?) and an ethical sense of responsibility (answering the question "Who is responsible for how it works?").

Principle-based approaches such as AI4People [Floridi *et al*., 2018], as well as recommendations issued by international organizations and regulatory bodies, have played an important role in consolidating a shared ethical vocabulary for AI. Initiatives such as the OECD (Organisation for Economic Co-operation and Development) Principles on Artificial Intelligence [OECD, 2019], UNESCO's Recommendation on the Ethics of Artificial Intelligence [UNESCO, 2021], and the European Commission's Ethics Guidelines for Trustworthy AI [European Commission, 2019] converge on core concerns, including transparency, accountability, human oversight, autonomy, and fairness. While these frameworks establish widely accepted normative expectations and support policy-making and governance efforts, they remain largely articulated at a high level of abstraction, offering limited guidance on how ethical principles should be concretely communicated and operationalized at the interface and interaction levels. [Morley *et al*., 2021; Shneiderman, 2020].

Ethical principles are often externalized in policies, terms of service, or compliance documents, which users rarely consult during everyday interactions. As a result, ethical commitments may exist formally without being meaningfully communicated during system use. This gap between abstract ethical intentions and concrete user experiences raises critical questions for HCI research, particularly regarding how ethical principles become visible, understandable, and actionable for users [Binns *et al*., 2018; Shneiderman, 2020].

As prior work has noted, the effectiveness of ethical principles does not depend solely on their formulation but on how they are enacted and perceived in practice [Morley *et al*., 2021]. Without deliberate design strategies that integrate ethical considerations into interaction flows, feedback mechanisms, and user controls, ethical principles risk remaining symbolic rather than operative. This limitation motivates the need to investigate ethics not only as a normative framework but also as a communicative phenomenon embedded in interactive systems. From an HCI perspective, this shift foregrounds the role of interfaces, interaction mechanisms, and system behaviors in mediating how ethical considerations become visible, intelligible, and actionable for users.

2.2 Semiotic Inspection Method (SIM) in Scientific Contexts

The SIM [de Souza *et al*., 2006; de Souza and Leitão, 2009] is grounded in Semiotic Engineering, an HCI theory that sees HCI as a form of computer-mediated communication between designers and users at interaction time [De Souza, 2005]. From this perspective, interaction is understood as a communicative process in which designers express their assumptions, intentions, and design rationale through the system interface. The interface conveys to users who the system is designed for, which goals and needs it addresses, and how interaction should occur to achieve the intended outcomes.

In Semiotic Engineering, this communication is mediated by signs, which are the means through which designers convey their intentions to users. Such signs include interface elements, messages, labels, controls, documentation, and system behaviors that users interpret while interacting with the system. Consequently, the interface with the system is composed of not only the visual representations displayed on the screen, but also of the signs that emerge through interaction itself, collectively communicating the designer's understanding of users, their goals, and the intended ways of using the system.

During use, users interpret the messages embedded in

the interface and gradually construct an understanding of the system based on the information made available through interaction. This communication process is enabled by the elements presented in the interface, which function as vehicles for meaning. Consequently, the interface can be seen as a meta-communication artifact: communication between designers and users occurs indirectly through users' interaction with the system itself.

According to Semiotic Engineering, this meta-communication can be summarized by the following template:

> *"Here is my understanding of who you are, what I've learned you want or need to do, in which preferred ways, and why. This is the system that I have designed for you, and this is how you can or should use it in order to fulfill a range of purposes consistent with this vision."* [De Souza, 2005, p.25]

To assess the quality of this designer-user communication, Semiotic Engineering introduces the concept of *communicability*, which refers to the extent to which a system effectively and efficiently conveys the designer's communicative intentions and the interaction principles underlying the design [De Souza, 2005]. When users are unable to correctly interpret this communication, *communication breakdowns* may occur, potentially hindering or even preventing the system from being used.

SIM is one of the evaluation methods proposed within Semiotic Engineering to examine communicability. It involves a systematic inspection of the system interface by an expert, whose task is to reconstruct the designer's intended meta-communication and identify points where communication may fail or become ambiguous [de Souza *et al*., 2006].

Before conducting the inspection, a preparatory phase is required. In this phase, the evaluator defines the objective of the inspection, determines its scope, and establishes a usage scenario that contextualizes the analysis. Following this preparation, SIM is carried out in five steps. The first three steps are iterative and focus on reconstructing the designer's meta-message from different perspectives, each corresponding to a specific type of interface sign [de Souza and Leitão, 2009]. These include *metalinguistic signs*, which explain or clarify other interface elements (such as error messages, tips, or help documentation); *static signs*, which represent the system state at a given moment (such as button labels, menu items, or toolbar icons); and *dynamic signs*, which reflect system behavior as it unfolds through user interaction (such as responses triggered by user actions). In the context of generative AI systems, dynamic signs include not only observable interaction behaviors but also textual outputs generated by the model during use. From the user's perspective, these outputs are part of the interface's communicative repertoire and therefore constitute legitimate objects of analysis, even when their production is only partially predictable or controllable by designers.

The application of this classification to generative AI systems introduces an additional interpretive challenge. Textual outputs produced by large language models exhibit characteristics of both static and dynamic signs. On the one hand, they are presented to users as textual artifacts that can be read and interpreted similarly to traditional static interface elements. On the other hand, they are generated during interaction in response to user inputs and may vary across sessions, contexts, and system versions. In this study, such outputs were operationally treated as dynamic signs because they emerge as part of the unfolding interaction and play a central role in shaping users' interpretation of the system. Nevertheless, their hybrid nature highlights a conceptual tension that deserves further investigation within Semiotic Engineering when applied to generative AI systems.

In the fourth step, the evaluator contrasts the interpretations derived from these three perspectives, examining inconsistencies and reflecting on how different interface elements were used to convey distinct aspects of the designer's message. In the fifth step, the evaluator consolidates these interpretations, producing an overall assessment of the system's communicability from the designer's standpoint.

SIM can also be applied in scientific research to support knowledge generation in HCI [de Souza *et al*., 2010]. In such contexts, two additional considerations are introduced. First, during the preparatory phase, the research question that the inspection aims to address must be explicitly defined. Second, a triangulation step is incorporated at the end of the analysis, in which findings are validated and strengthened through comparison with results obtained by other experts or through complementary methods.

Prior studies have applied SIM in scientific contexts to investigate the future impacts of choices about how digital legacies will be handled [Prates *et al*., 2015; Pereira *et al*., 2016], the communicative strategies employed by chatbots to inform users about their characteristics and functionalities [Valério *et al*., 2017], and the ethical aspects of generative AI systems [Gomes *et al*., 2025].

Specifically, in the context of generative AI systems, SIM offers distinct advantages. It allows researchers to analyze how ethical principles are communicated without requiring access to proprietary model internals, focusing instead on what users can perceive and interpret through interaction [Gomes *et al*., 2025]. However, the application of SIM to generative AI systems also raises important methodological and theoretical challenges. Unlike conventional interactive systems, large language models are capable of producing signs dynamically and non-deterministically, generating outputs that may not be fully predictable or directly controlled by their designers. Consequently, the relationship between the designers' intended meta-message and the communicative meanings emerging during interaction becomes less stable and more difficult to reconstruct.

Rather than assuming a perfect correspondence between designer intentions and system outputs, SIM enables the investigation of how ethical meanings are manifested and interpreted through the signs available to users at a particular moment in time. In this sense, the method supports the analysis of ethical communicability even when the origins of specific outputs remain partially opaque. Moreover, because the same inspection protocol can be systematically applied across different versions of a system, SIM provides a useful lens for examining how communicative patterns, ethical cues, and potential communicability breakdowns evolve over time. These characteristics make SIM a suitable method-

ological choice for investigating ethical communicability in generative AI systems, while also highlighting important open questions regarding authorship, designer agency, and the stability of meta-communication in non-deterministic interactive systems.

2.3 Ethical Communication in Interactive Systems

From an HCI perspective, ethics can be understood as a matter of communicability, that is, how systems convey values, responsibilities, risks, and limitations to users through interaction. As discussed in the previous subsection, Semiotic Engineering conceptualizes interaction as a communicative process in which designers express assumptions and intentions through the interface. Building on this view, ethical communication concerns how such intentions and values are made intelligible to users during system use. When ethical considerations are clearly conveyed, users are better able to form appropriate mental models of system capabilities and constraints; when ethical information is hidden, fragmented, or ambiguous, informed use and trust may be undermined [Shneiderman, 2020; Binns *et al*., 2018].

In interactive systems, ethical considerations are communicated through multiple interface elements and interaction mechanisms. Explicit information provided through policies, disclaimers, and help texts outlines system commitments, limitations, and responsibilities, while interface components such as labels, icons, settings, and controls indicate available actions and how user data and inputs are managed. In addition, system behaviors observed during interactions, including refusals, warnings, explanations, and feedback, play a particularly important role in shaping users' understanding in situated use, especially in ethically sensitive scenarios where system limitations, risks, or responsibilities become salient [de Souza and Leitão, 2009].

To support systematic ethical reflection on such communicative phenomena, Barbosa *et al*. [2021] propose a semiotics-based epistemic tool for reflecting on ethical issues related to the design of digital technologies in general, and AI in particular [Barbosa *et al*., 2024; Nunes *et al*., 2024]. The core element of this tool is the *Extended Metacommunication Template* (EMT), which extends Semiotic Engineering's original metacommunication template by incorporating ethical questions associated with different stages of the system lifecycle, including analysis, design, implementation, deployment, and post-deployment use. Rather than focusing solely on what designers communicate to users through the interface, the EMT supports reflection on the ethical assumptions, responsibilities, consequences, and trade-offs embedded in this communication.

Figure 1 illustrates the EMT, showing how the stages of the design and development lifecycle, the metacommunication template from Semiotic Engineering, and the ethical questions that must be answered in each stage are related.

In the context of generative AI systems, this epistemic tool is particularly relevant. The non-deterministic and conversational nature of these systems introduces additional challenges for ethical communication, as system behavior may vary across interactions, prompts, and versions [Bender *et al*., 2021]. Such variability complicates the communication of stable expectations regarding capabilities, limitations, and risks, and makes ethical responsibilities harder to locate. By structuring ethical reflection around lifecycle stages and explicit guiding questions, the epistemic tool supports the interpretation of how ethical responsibilities are distributed, negotiated, or obscured in evolving AI systems.

Empirical studies have shown that ethical communication in AI systems is often uneven and principle-specific [Gomes *et al*., 2025]. Some ethical principles tend to be addressed through explicit interface mechanisms, while others remain largely implicit or externalized. Interpreted through the combined lens of the SIM and the epistemic tool proposed by Barbosa *et al*. [2021], these asymmetries suggest that ethical communicability should be understood as a matter of degree and consistency rather than as a binary property. This perspective reinforces the relevance of analytical approaches capable of systematically comparing how ethical meanings and responsibilities are communicated across systems, principles, interaction contexts, and system versions.

3 Methodology

This section presents the methodological approach used to analyze how ethical communicability in generative AI systems evolves across successive system versions.

3.1 Research Design

This study adopts a qualitative, longitudinal, and comparative research design to investigate how ethical communicability in generative AI systems evolves over time by systematically comparing successive versions of the same systems.

The extension is grounded in the assumption that generative AI systems are dynamic interactive artifacts that undergo frequent updates, including interface redesigns, policy changes, and adjustments to underlying models. Such changes may affect how ethical principles are communicated to users.

Importantly, the longitudinal comparison does not target specific underlying model versions in isolation. Rather, the unit of analysis is the publicly available version of each generative AI system as experienced by users during the inspection periods. Consequently, the observed changes may result from a combination of modifications to foundation models, guardrails, interface elements, interaction flows, policies, and governance mechanisms. From the perspective of ethical communicability, the focus is not on identifying the technical source of each change, but on examining how such changes become visible through the signs available to users and influence the communication of ethical considerations during interaction.

By adopting a longitudinal perspective, the present study seeks to identify patterns of continuity, improvement, regression, and fragmentation in ethical communication across system versions.

Methodological continuity with the previous study was deliberately preserved to ensure comparability. The same analytical framework, ethical principles, inspection scenarios, and core methodological procedures were retained, while the scope of analysis was expanded to include temporal comparison and cross-version interpretation.

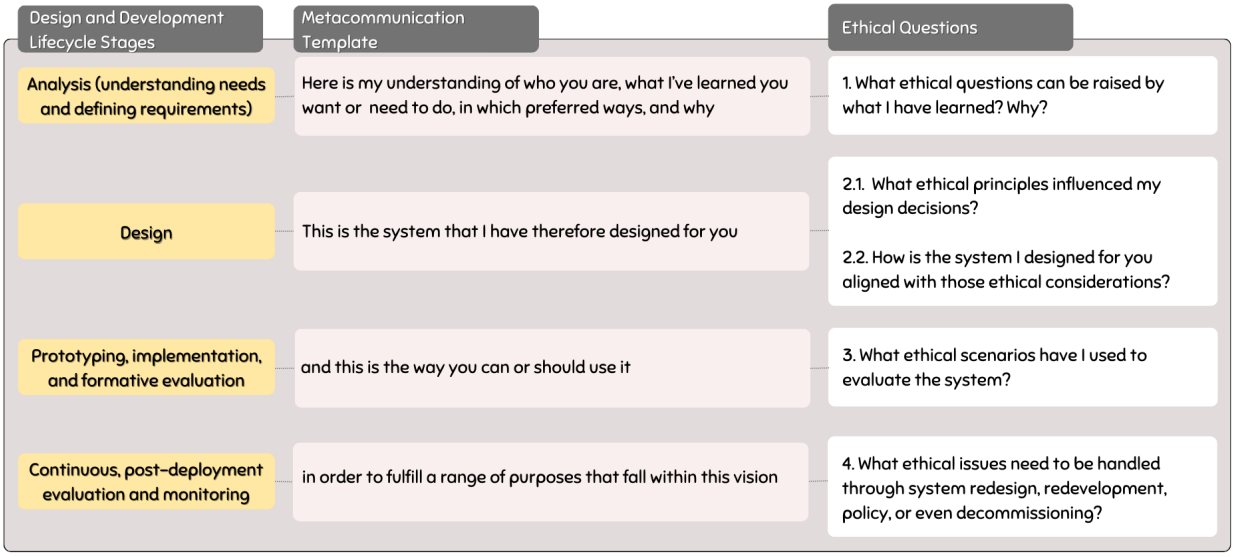


Figure 1. Alignment of the design and development lifecycle stages with the meta-communication template and ethical questions. Adapted from Barbosa *et al*. [2021]

3.2 Selection of Generative AI Systems

The generative AI systems selected were the same as those analyzed in [Gomes *et al*., 2025], namely ChatGPT, Gemini, and Claude, with the aim of preserving analytical consistency while enabling longitudinal comparison. In the mentioned study, systems were selected based on the following criteria: (i) widespread public use; (ii) availability through publicly accessible interfaces; (iii) support for conversational interaction with text-based generation; and (iv) relevance to socially significant contexts such as information seeking, creative work, and decision support.

For each system, the version analyzed in Gomes *et al*. [2025] was revisited and compared with its most recent publicly available version at the time of the extended study. In this context, the term "version" refers to the publicly available system configuration experienced by users, encompassing interface elements, interaction flows, policies, safeguards, and underlying model updates whenever these changes are reflected in the user experience. Since providers of commercial generative AI do not always disclose all technical modifications introduced over time, the comparison focuses on observable changes in ethical communicability rather than on specific model releases or internal architectural updates. The analysis, therefore, considers changes in interface elements, interaction flows, system feedback, disclaimers, controls, and other signs relevant to ethical communication, regardless of whether they originate from model updates, guardrail adjustments, policy revisions, or interface redesigns.

To support transparency and reproducibility, the inspection scope for each system was documented, including version identifiers when available, inspection dates, and the specific interface features analyzed. The analysis focuses exclusively on functionalities considered in the previous study, i.e., text generation accessible in free or default configurations, reflecting typical user experiences and avoiding the influence of premium or experimental features.

3.3 Ethical Principles and Ethically Framed Application of the Semiotic Inspection Method

The analysis was guided by the ethical principles articulated in the AI4People framework [Floridi *et al*., 2018] that include Beneficence (B), Non-Maleficence (nM), Autonomy (A), Justice (J), and Explicability (E), which provide a normative reference for examining how ethical considerations are communicated through generative AI interfaces. Rather than treating these principles as abstract values or external evaluation criteria, this study operationalizes them as analytical lenses that inform and structure the inspection process itself.

The SIM was applied as a scientific inspection method explicitly framed by ethical reflection. Rather than using the SIM solely to evaluate general communicability, its application was oriented toward examining how ethical considerations are encoded, communicated, and potentially broken down at the interface level of generative AI systems, aligning the method with the study's focus on ethical communicability. For this purpose, as illustrated in Figure 2, the ethical questions proposed by Barbosa *et al*. [2021] (shown in Figure 1) were integrated throughout the SIM process, guiding all steps from the creation of the interaction scenario to the reconstruction of the meta-messages and the interpretation of communication breakdowns, serving both as analytical lenses and as outcomes of the inspection. This integration ensured that ethical reflection was not treated as a post hoc

layer but as an integral component of the inspection process itself.

Although Figure 2 associates specific ethical questions with particular stages of the inspection, these questions were not treated as isolated checkpoints or independent analyses. Rather, they were applied to the reconstructed metacommunication message as a whole. The first question of the Extended Metacommunication Template (*What ethical questions can be raides by what I have learned? Why?*) supported the identification of ethical issues emerging from the meta-message, while the subsequent questions guided the analysis of ethical principles and communicability breakdowns throughout the inspection process.

In the preparatory stage, we defined the guiding research question to be answered by applying the SIM method as follows: *How do designers of generative AI systems communicate ethical considerations to users through interface elements and system behavior, and how does this communication evolve over time?*

Also, in the SIM preparation stage, we took the same ethical inspection scenarios adopted in the IHC 2025 study [Gomes *et al*., 2025] (ethical question 3 in Figure 2) to elicit ethically relevant system behaviors during interaction [1]. The scenarios reflect realistic use situations in which ethical principles become salient, such as requests involving sensitive information, potentially harmful content, biased outputs, or unclear system limitations. They were designed to avoid intentionally generating harmful content, focusing instead on how systems communicate risks, constraints, responsibilities, and safeguards. By maintaining the same inspection scenarios across system versions, the study ensures methodological consistency and comparability between inspection rounds, allowing observed differences in ethical communicability to be attributed to system evolution rather than variations in interaction context. This consistency is essential for supporting a valid longitudinal comparison. Although the reuse of identical inspection scenarios increases comparability across inspection rounds, the stochastic nature of generative AI systems means that identical prompts may still produce different outputs. Consequently, the study does not seek to attribute observed differences to specific technical updates, but rather to compare patterns of ethical communicability manifested during the respective inspection periods.

Once the activities of the SIM preparation stage were concluded, the inspection of each system followed the canonical SIM steps, encompassing the analysis of metalinguistic, static, and dynamic interface signs. Metalinguistic signs (e.g., policies, disclaimers, and help documentation) were inspected to identify how ethical commitments, responsibilities, and limitations are explicitly articulated. Static signs (e.g., labels, settings, and controls) were analyzed as indicators of how ethical principles are represented in the interface. Dynamic signs included not only interaction behaviors designed by the platform but also signs generated by the AI model itself, such as responses, refusals, warnings, explanations, and feedback produced during interaction. outputs were analyzed as dynamic signs because they constitute part of the communicative experience available to users and frequently make ethical considerations salient in non-deterministic and conversational contexts. Although such signs may not always be fully predictable or directly controlled by designers, they nevertheless contribute to the ethical meanings communicated through the system and were therefore considered within the scope of the inspection.

During the inspection, the SIM was used to reconstruct the complete metacommunication message conveyed by the system through metalinguistic, static, and dynamic signs. The ethical questions proposed by Barbosa *et al*. [2021] were incorporated throughout the analysis as a framework for ethical reflection on the reconstructed meta-message. More specifically, the application of the first question of the Extended Metacommunication Template ("What ethical questions can be raised by what I have learned? Why?") led to the identification of ethical issues related to assumptions about users, intended uses, system capabilities, limitations, responsibilities, and potential impacts (ethical question 1 in Figure 2). The resulting ethical questions then supported the analysis of how the systems addressed the principles of Beneficence, Non-Maleficence, Autonomy, Justice, and Explicability (ethical questions 2.1 and 2.2 in Figure 2).

It is important to note that the ethical questions were not treated as evaluation tasks to be directly answered by the inspectors. Rather, they emerged from the application of the first EMT question ("What ethical questions can be raised by what I have learned? Why?") to the reconstructed metacommunication message. These questions then functioned as analytical prompts that guided the search for communicative evidence related to ethical principles, such as warnings, explanations, controls, refusals, policies, and other interface signs. For example, a question concerning energy consumption did not require estimating the system's actual computational costs; instead, it motivated the analysis of whether and how the system communicated responsibilities, impacts, or limitations associated with its operation.

Finally, in the last step of the SIM, we evaluated the communicability of the system, highlighting communicability breakdowns and ethical issues associated with one or more ethical principles (ethical question 4 in Figure 2).

To support the study's longitudinal perspective, the SIM results were compared across different versions of each generative AI system. The first round of inspections took place between July and December 2024. The second round of inspections, with the updated versions of the systems, took place between November and December 2025. This comparison made it possible to identify changes in the communication of ethical considerations over time, highlighting how ethical communicability evolves alongside system updates. The method thus supported a comparative analysis not only across systems but also across temporal snapshots of the same system, making it possible to examine ethical communicability as an evolving property of interactive systems rather than as a static design attribute.

In both rounds of inspections, a two-stage triangulation was employed to strengthen the scientific validity of the findings. First, inspections conducted independently by two researchers were compared and consolidated through discussion, with the support of a SIM expert when necessary. Second, consolidated results were triangulated across systems to

[1] Inspection scenarios are available in Appendix A

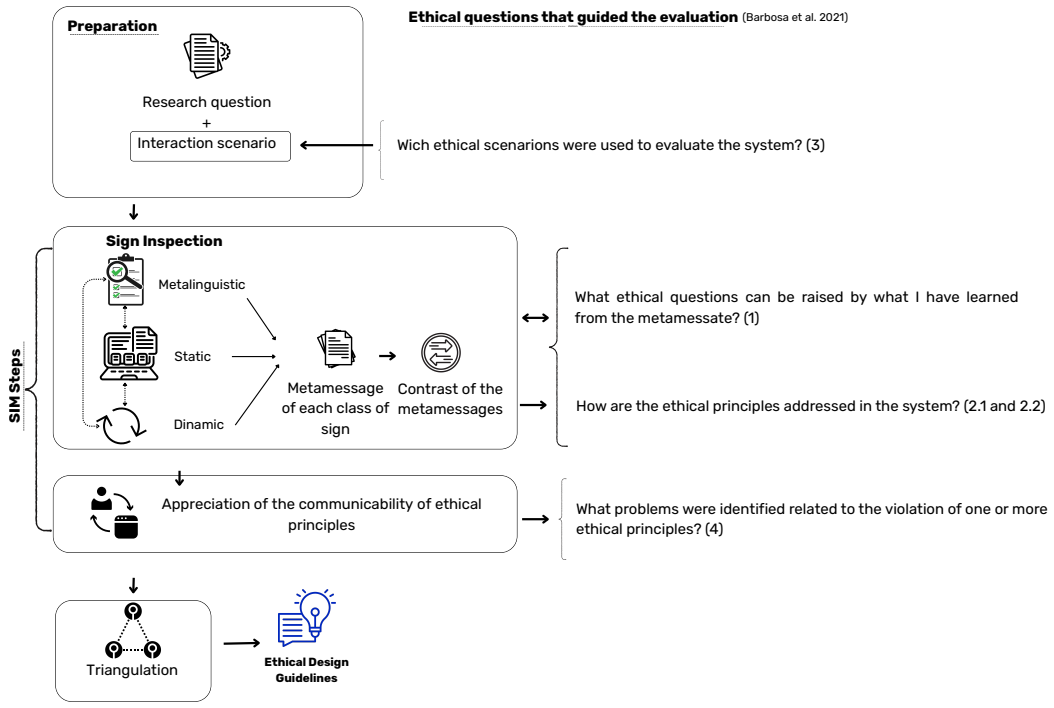


Figure 2. Overview of the SIM-based inspection process, showing how the ethical questions of the Extended Metacommunication Template (EMT) were applied to the reconstructed metacommunication message to identify ethical issues, analyze ethical principles, and evaluate communicability breakdowns.

identify recurring patterns, divergences, and gaps in ethical communication. This triangulation reinforced the robustness of the analysis and supported the generation of transferable insights regarding ethical communicability in generative AI systems.

3.4 Ethical Considerations

This study did not involve the collection of data from human participants and, therefore, did not require approval from a Research Ethics Review Board. The research consisted of a systematic inspection of publicly available generative AI systems, focusing on their interface elements, documentation, and interactional behaviors. All analyzes were conducted exclusively by the authors, applying the SIM in a scientific and non-intrusive manner.

Despite the absence of human subjects, ethical considerations were addressed throughout all stages of the research, from planning to analysis and reporting. A non-intrusive methodology was adopted, avoiding the collection of personal data, interference with users, or deliberate circumvention of system safeguards, while inspection scenarios were designed to reflect realistic and ethically relevant situations.

Some inspection scenarios involved ethically sensitive requests intended to examine how systems communicate risks, limitations, and safeguards. These scenarios were plausible but fictional situations created exclusively for research purposes and did not correspond to real-world cases or decision-making contexts. Any outputs generated during these inspections were analyzed solely for research purposes and were not used, shared, or disseminated beyond the scope of the study. The objective was not to produce or promote harmful content, but to understand how the systems respond to ethically relevant situations.

The study followed principles of transparency, reproducibility, and responsible research practices. System selection was guided by relevance and representativeness, focusing on widely used generative AI systems with publicly accessible interfaces. A consistent and documented analytical protocol was applied across systems and versions, enabling comparability and supporting replicability. Analytical claims were grounded strictly in observable interface signs and documented behaviors, avoiding speculative interpretations or attributions of intentionality beyond what could be inferred from the systems' communicative cues.

The research is aligned with established ethical standards for scientific publication and conduct, including the SBC Code of Conduct and the principles of the Committee on Publication Ethics (COPE). These guidelines informed decisions related to integrity, accountability, proper attribution, and transparency throughout the study.

Finally, broader societal and environmental implications of generative AI are explicitly recognized as part of the ethical responsibility of conducting and reporting research on AI-based interactive systems. This concern is particularly relevant for contemporary systems powered by large language models (LLMs), whose large-scale computational requirements have raised growing concerns regarding energy consumption, resource use, and societal impacts.

4 Results: Longitudinal Analysis of Ethical Communicability

This section presents the results of the longitudinal analysis of ethical communicability in the three generative AI systems, comparing their 2024 and 2025 versions. First, Section 4.1 outlines the ethical questions that guided the inspections, grounding the analysis in concrete interactional concerns. Section 4.2 then provides an overview of the main changes observed across system versions, highlighting cross-system trends and divergences. Building on this overview, Section 4.3 examines the evolution of each AI4People ethical principle over time. Finally, Section 4.4 synthesizes cross-cutting patterns and inconsistencies, discussing fragmentation across principles, tensions between metalinguistic commitments and interactional cues, and uneven trajectories of ethical communicability across systems.

4.1 Ethical Questions Guiding the Inspection of the Systems

To guide the inspection of the new versions of the AI generative systems from the perspective of addressing ethical principles, the same ethical questions from the original study were retained to ensure comparability. They are listed below (in parentheses are the ethical principles to which the questions refer):

- What if the system consumes many energy resources for its processing? *(B)*
- What if the user asks the system to perform an inappropriate task, for example, generating false or harmful information about one or more individuals? *(B, nM)*
- What if the user tries to circumvent the system's guidelines in order to obtain information that contradicts established policies? *(B, nM)*
- What if the system collects user data and/or shares it with other platforms without consent? *(nM)*
- What if the user is not familiar with this type of tool and cannot write commands correctly, clearly, and concisely, thus obtaining incorrect results? *(nM, A)*
- What if the system does not allow the user to make decisions and control the path of interaction to be followed? *(A)*
- What if the user has some type of disability that prevents them from interacting with the system in a conventional way? *(J)*
- What if the system creates inequalities in access to information for users with different educational and socioeconomic backgrounds? *(J)*
- What if the system is difficult to use and does not clarify the basis on which the generated information relies? *(J, E)*

4.2 Overview of Changes Across System Versions

The comparative analysis between the 2024 and 2025 versions of the ChatGPT, Gemini, and Claude systems reveals a meaningful reconfiguration in how the ethical principles of Beneficence, Non-Maleficence, Autonomy, Justice, and Explainability are treated and communicated to users. In the 2024 versions, these principles were present mainly in a diffuse and fragmented way, anchored primarily in documents such as terms of use and privacy policies, and poorly incorporated into the interactive experience itself. Ethics, in this context, operated more as legal and discursive support than as an element effectively integrated into the interaction design, which generated ambiguities, informational gaps, and ruptures in ethical communicability, as discussed in [Gomes *et al*., 2025].

In general, previously, the systems communicated benefits in an abstract way, generically warned about risks, transferred much of the responsibility for appropriate use to the user, and offered superficial explanations about their limitations and operation. Protection against harm, respect for user autonomy, and concerns related to social justice appeared more as stated intentions than as clear guidelines for responsible use. The absence of contextualized explanations and consistent ethical cues throughout the interaction meant that users had to infer when and how to trust the system, which weakened both informed decision-making and the perception of institutional responsibility.

In the 2025 versions, an evolution towards a more explicit and operational treatment of some ethical principles, particularly explicability and non-maleficence, is observed, now communicated more explicitly and integrated into the interface. Instead of remaining restricted to lengthy legal texts, ethical values are reiterated through contextual warnings, explanations of capabilities and limitations, clearer restrictions on potentially harmful uses, and messages that encourage human oversight and external verification. This change indicates a shift from ethics as a mere normative statement to ethics as a guiding element of interaction, approaching the notion of tangible requirements proposed by Morley *et al*. [2021]. In comparative terms, the "before" can be characterized by a reactive and less visible ethical stance focused on mitigating legal risk, whereas the "now" reflects a more proactive communicational ethic oriented toward reducing ambiguity by explicitly stating expected benefits, usage limits, associated risks, and potential impacts.

Despite these advances, changes across system versions are neither uniform nor holistic. Additions are concentrated mainly on explicability-related cues and harm prevention mechanisms, while refinements dominate over radical redesigns. Refinements typically involve adjustments in wording, tone, timing, or placement of existing ethical messages, improving consistency but sometimes reducing contextual richness.

Table 1 synthesizes the cross-system comparison between the 2024 and 2025 versions of ChatGPT, Gemini, and Claude, highlighting how additions and refinements in ethical communicability are distributed across ethical principles and systems. Rather than indicating a uniform progression, the table reveals principle-specific and system-specific trajectories, showing that explicability consistently receives more explicit communicative support across all systems, while beneficence and justice remain largely implicit. By organizing these changes comparatively, the table supports the interpretation of ethical communicability as an evolving yet uneven property of generative AI interfaces, shaped by incremental and non-uniform design decisions across system

versions. These differentiated trajectories motivate a more detailed examination of the results. In the next section, we therefore focus on each ethical principle individually, analyzing how Beneficence, Non-Maleficence, Autonomy, Justice, and Explicability evolve over time across the analyzed systems.

4.3 Evolution of Ethical Principles Over Time

This section examines how the ethical principles articulated by the AI4People framework evolve between the 2024 and 2025 versions of the analyzed systems. Rather than treating ethics as a static property, the analysis highlights continuities, improvements, and areas of stagnation in ethical communicability over time.

Beneficence Across systems, beneficence remains the least explicitly operationalized of the ethical principles. In the 2024 versions, benefits were primarily communicated through abstract institutional statements, framing the systems as generally helpful or socially beneficial, without clear articulation at the interface level. This pattern largely persists in 2025, indicating continuity rather than structural change.

Nevertheless, a limited improvement is observable. In the newer versions, systems increasingly emphasize tangible and immediate benefits in specific interactional contexts, particularly when handling sensitive data or supporting user decision-making. As illustrated in Figure 3, benefits are more frequently associated with concrete functionalities and situational guidance rather than broad normative claims. Despite this refinement, beneficence remains mostly implicit and weakly integrated into interaction design. An additional observation is that communicative mechanisms associated with beneficence remain predominantly conveyed through metalinguistic signs, such as introductory messages, feature descriptions, and institutional discourse. Unlike principles such as Non-Maleficence and Explicability, which increasingly appear through dynamic interactional mechanisms, beneficence continues to be communicated primarily through explanatory content rather than through situated interaction. This pattern has remained largely stable across versions, suggesting that the operationalization of Beneficence at the interface level has evolved more slowly than other ethical principles.

Non-Maleficence Non-maleficence shows the most consistent evolution between 2024 and 2025. Earlier versions relied primarily on generic disclaimers about possible errors or misuse, often transferring responsibility entirely to users. This created a tension between cautionary discourse and assertive system behavior.

In 2025, harm-prevention mechanisms become more explicit and interactionally embedded. Systems increasingly refuse to engage in sensitive or potentially harmful requests[2], and provide clearer warnings about misinformation risks during use[3]. Beyond warning users about the possibility of inaccurate information, Claude also provides guidance on exploring techniques to minimize hallucinations and ensure accurate and trustworthy outputs. This behavior represents a more proactive communicative strategy, combining risk disclosure with actionable recommendations to support safer and more informed use. Gemini further refines its communicative stance through changes in message tone between versions (Figure 4). These changes indicate a clear improvement, reducing misalignment between ethical commitments and system behavior, even though some residual risks remain.

Autonomy Autonomy evolves through incremental refinements rather than radical redesign. In 2024, users were formally informed about data-sharing choices, but control mechanisms were often opaque or poorly integrated into the interaction flow, limiting informed decision-making.

In 2025, autonomy is more visibly supported through interface-level controls. Users gain clearer options to manage data and interaction histories, such as temporary chats (Figure 5) and privacy-related settings addressing data protection and misuse prevention[4]. ChatGPT further strengthens mediated autonomy through parental control mechanisms (Figure 6), extending user governance to vulnerable audiences. These developments represent a moderate refinement; however, explanations of how such choices concretely affect system behavior remain limited.

Justice Justice remains largely implicit across both versions, revealing relative stagnation. In 2024, references to fairness and accessibility appeared mainly as acknowledgments of potential bias, without clear interface-level strategies for inclusion or non-discrimination.

In the 2025 versions, some policy-level refinements are observable, such as stronger alignment with regulations, parental controls, and statements about equitable use. However, these changes rarely translate into concrete, interface-level communicative cues. Thus, overall, justice exhibits relative stagnation, with minor refinements but no substantial shift toward explicit, interaction-embedded communication.

Explicability Explicability exhibits the strongest and most visible improvement over time. In 2024, explanations about system operation, data use, and limitations were fragmented and often externalized to documentation, creating cognitive asymmetries during interaction.

In 2025, systems provide clearer and more accessible explanations about how content is generated and how data is used. Gemini explicitly clarifies aspects of model training[5], while ChatGPT provides more detailed explanations regarding data usage[6]. Claude further reinforces ex-

[2]Figure available in the external repository `https://github.com/Libianegomes/ethical-ai-longitudinal-SIM-analysis/blob/main/non-maleficence/ChatGPT_Gemini_2025_nonmaleficence1.pdf`. Accessed on 15 July 2026.

[3]Figure available in the external repository `https://github.com/Libianegomes/ethical-ai-longitudinal-SIM-analysis/blob/main/non-maleficence/Chatgpt_Claude_2025_nonmaleficence2.pdf`. Accessed on 15 July 2026.

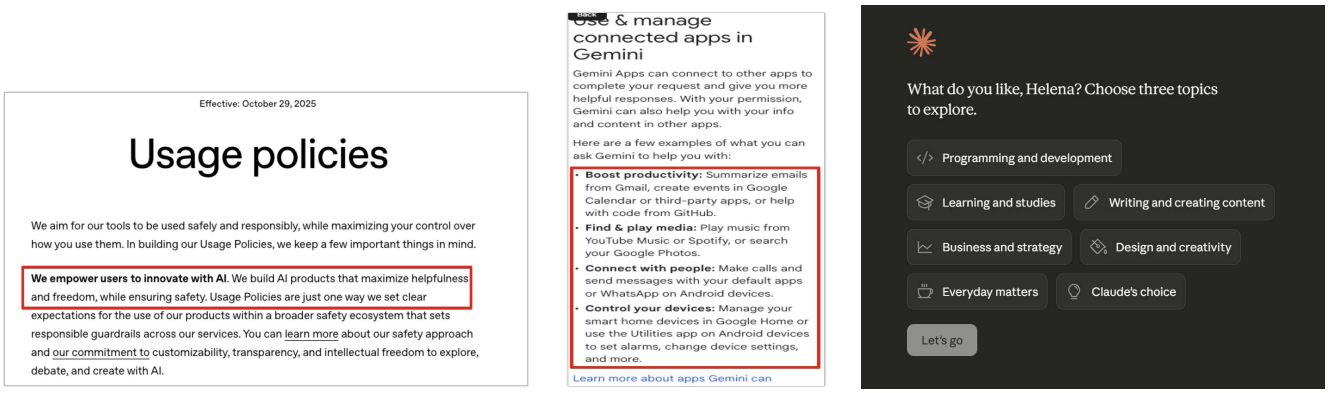
[4]Figure available in the external repository `https://github.com/Libianegomes/ethical-ai-longitudinal-SIM-analysis/blob/main/autonomy/ChatGPT_Gemini_Claude_2025_autonomy1.pdf`. Accessed on 15 July 2026.

[5]Figure available in the external repository `https://github.com/Libianegomes/ethical-ai-longitudinal-SIM-analysis/blob/main/explicability/Gemini_2025_explicability1.pdf`. Accessed on 15 July 2026.

[6]Figure available in the external repository `https://github.`

Table 1. Cross-system comparison of changes in ethical communicability between the 2024 and 2025 versions of ChatGPT, Gemini, and Claude.

Ethical Principle	ChatGPT (2024 → 2025)	Gemini (2024 → 2025)	Claude (2024 → 2025)
Beneficence	Maintain largely implicit; Increased emphasis on tangible user benefits	Maintain largely implicit; Increased emphasis on tangible user benefits	Maintain largely implicit; Increased emphasis on the direct and positive impacts on users' everyday activities
Non-Maleficence	More explicit harm-prevention mechanisms; Greater user awareness of the risks	Clearer signaling of potential inaccuracies, reinforcing harm prevention	More explicit communication of robust safety and data-protection measures
Autonomy	Stronger preventive controls - particularly regarding data sharing and vulnerable users	Clearer privacy controls and data opt-out options	Stronger emphasis on personalization, enabling users to tailor aspects of the interaction to their preferences and needs
Justice	Maintain largely implicit; Stronger focus on policies aimed at responsible and fair use	Maintain largely implicit; no addition	Maintain largely implicit; Clearer policy-driven strengthening, with explicit commitments to equity, accessibility, and non-discrimination
Explicability	More explicit explanations of content generation and data use	Clearer explanations regarding data usage and the AI training process	Clearer strengthening through the addition of tutorials, feedback, and more explicit communication about errors



(a) ChatGPT **(b)** Gemini **(c)** Claude

Figure 3. Informing the user the systems benefices

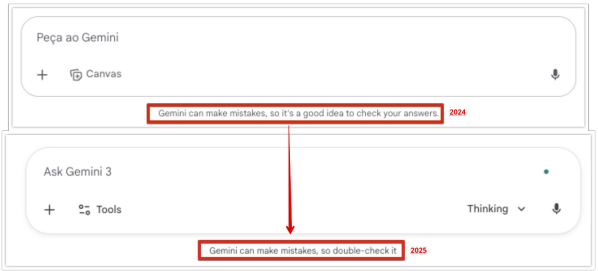


Figure 4. Change in message tone to check the content generated between Gemini versions 2024 and 2025

plicability through tutorials and structured feedback mechanisms[7]. Although deep technical explainability remains limited, these additions significantly enhance discursive transparency and position explicability as the most communicatively supported ethical principle across systems.

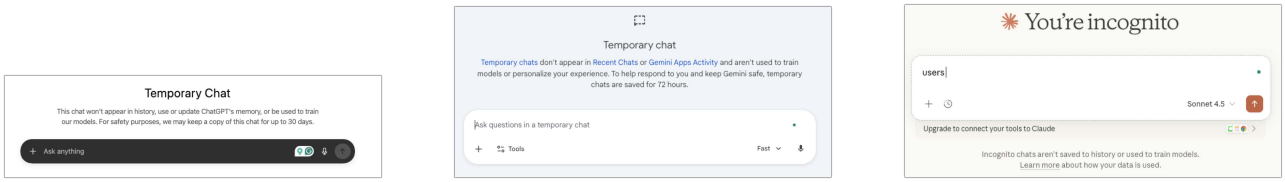
Taken together, the longitudinal analysis reveals that ethical principles do not evolve uniformly. While non-maleficence and explicability show clear communicative gains, beneficence and justice remain largely implicit, and autonomy advances unevenly through incremental controls. This uneven evolution reinforces the argument that ethical communicability in generative AI systems is an emergent, principle-specific, and non-linear property, shaped by selective design priorities rather than holistic ethical strategies. In the next section, we discuss the implications of these findings for HCI research and for the design and evaluation of ethically communicative AI systems.

4.4 Cross-Cutting Patterns and Inconsistencies

Across systems and versions, the results reveal a set of recurring cross-cutting patterns that qualify the observed evolution of ethical communicability. Rather than indicating a coherent and holistic ethical maturation, the changes between 2024 and 2025 expose structural fragmentation across ethi-

com/Libianegomes/ethical-ai-longitudinal-SIM-analysis/blob/main/explicability/ChatGPT_2025_explicability2.pdf. Accessed on 15 July 2026.

[7]Figure available in an external repository https://github.com/Libianegomes/ethical-ai-longitudinal-SIM-analysis/blob/main/explicability/Claude_2025_explicability3.pdf. Accessed on 15 July 2026.



(a) ChatGPT **(b)** Gemini **(c)** Claude

Figure 5. Temporary chats

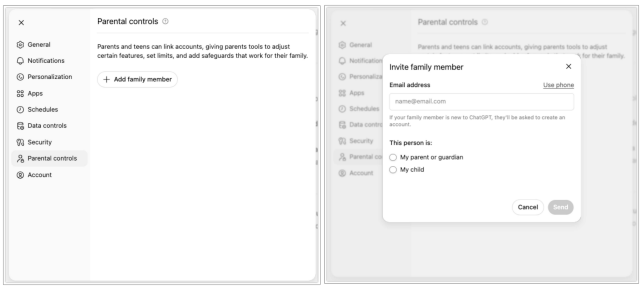


Figure 6. ChatGPT parental control function

cal principles, persistent tensions between discursive commitments and interactional cues, and uneven trajectories of evolution across systems.

A first recurring pattern is the *fragmentation across ethical principles*. Ethical communicability does not evolve uniformly across the AI4People principles. Explicability and non-maleficence receive the most consistent attention, with clearer warnings, refusals, explanations, and safety-related cues increasingly embedded in interaction. In contrast, beneficence and justice remain weakly articulated and largely implicit, often confined to abstract institutional statements or policy-level commitments. Autonomy occupies an intermediate position: while new controls and settings enhance user agency, their implications for system behavior are not always clearly communicated. This fragmentation suggests that ethical principles are treated as loosely coupled concerns rather than as an integrated ethical stance communicated through the interface as a whole.

A second cross-cutting issue concerns the *tension between metalinguistic commitments and interactional cues*. In both 2024 and 2025, ethical intentions are frequently articulated through metalinguistic signs, such as policies, help centers, and institutional statements, that express responsibility, transparency, or concern for user well-being. However, these commitments are not always consistently reinforced by static and dynamic signs encountered during everyday interaction. Although the 2025 versions show improvements, especially through contextual warnings, refusals, and explanations, misalignments persist. In some cases, interactional behavior continues to show confidence and fluency even when metalinguistic signs emphasize uncertainty, limitations, or risk, reproducing communicability breakdowns previously identified in the previous study.

Finally, the analysis highlights an *uneven evolution across systems*. While all three systems exhibit incremental changes, their trajectories differ in focus and consistency. Claude shows a more systematic consolidation of safety- and transparency-related communication, Gemini advances more consistently in the area of informational autonomy, offering detailed explanations about data collection, use, retention, and deletion, as well as explicit controls for managing activity, making it more transparent what is shared and how this can be controlled by the user, and ChatGPT incorporates clearer governance mechanisms for its use, such as parental controls and recurring messages reinforcing the need for human supervision and external review, which strengthens mediated autonomy and the protection of vulnerable audiences. These differences indicate that ethical communicability evolves through system-specific design decisions rather than through a shared or standardized approach to ethical communication in generative AI. As a result, users encounter divergent ethical messages and expectations depending on the system, reinforcing the notion that ethical communicability remains contingent, partial, and unevenly operationalized.

Taken together, these cross-cutting patterns reinforce the interpretation of ethical communicability as an evolving but unstable property of generative AI interfaces. Improvements between 2024 and 2025 coexist with persistent gaps, inconsistencies, and trade-offs, underscoring the need for more integrated, interaction-centered, and principle-spanning approaches to communicating ethics in generative AI systems.

5 Discussion

This section discusses the implications of the longitudinal findings presented in Section 4, moving beyond descriptive results to reflect on their conceptual, methodological, and design-oriented significance for HCI. By examining how ethical communicability evolves across system versions, the discussion situates ethical communication as a dynamic property of generative AI interfaces, shaped by ongoing design decisions rather than static statements. The section is structured around four complementary perspectives: ethical communicability as an evolving design property, implications for interface design, methodological reflections on the use of the SIM in longitudinal contexts, and broader ethical and societal considerations of generative AI systems.

5.1 Ethical Communicability as an Evolving Design Property

The results of this extended study reinforce the argument that ethical communicability in generative AI systems should be understood as an evolving phenomenon rather than as a stable property of a particular system version. Although any inspection necessarily captures a specific moment in the evolution of a system, the longitudinal perspective adopted here makes it possible to identify patterns of continuity, refinement, and change that would not be visible through the analysis of a single version alone. This is particularly important in the context

of generative AI, where systems evolve rapidly in response to technical developments, user practices, emerging societal concerns, and regulatory pressures.

This shift from a single-version perspective to a longitudinal one aligns with the semiotic understanding of interaction as an ongoing communicative process between designers and users, mediated by interface signs [De Souza, 2005]. Rather than treating ethical communicability as a fixed property of an interface, the results show that it evolves as designers revise the meta-communication encoded in the system over time. Incremental changes in wording, placement, or timing of ethical cues, rather than radical redesign, play a central role in shaping users' interpretation of responsibilities, risks, and system limitations. From this perspective, ethical communicability emerges as a temporally situated phenomenon whose quality depends on how communicative intentions are sustained, revised, or weakened across successive versions of the system.

The longitudinal findings further highlight that the evolution of ethical communicability is uneven across ethical principles and across systems. Some principles, particularly explicability and non-maleficence, show clearer communicational gains between 2024 and 2025, becoming more operationalized through contextual warnings, refusals, and explanations embedded in interaction. These changes align with principle-based approaches to responsible AI, which emphasize transparency, risk mitigation, and accountability as central ethical requirements [Floridi *et al*., 2018; Morley *et al*., 2021]. In contrast, beneficence and justice remain largely implicit, confined to abstract institutional discourse or policy-level commitments, with limited translation into concrete interface-level cues. Autonomy occupies an intermediate position, advancing through the introduction of controls and settings, yet often without sufficient explanation of their implications for system behavior.

Part of this asymmetry can be observed in the types of signs through which different ethical principles are communicated. While Non-Maleficence and Explicability increasingly rely on dynamic and contextualized interactional mechanisms, such as warnings, refusals, and explanations embedded in user interactions, Beneficence remains predominantly communicated through metalinguistic signs, including feature descriptions, introductory messages, and institutional discourse. This suggests that, despite modest refinements over time, Beneficence continues to be expressed primarily at a declarative level rather than being operationalized through situated interaction. As a result, its communicability appears less mature and less integrated into the user experience than that of principles more directly supported by interactional mechanisms.

One possible explanation for this uneven evolution lies in the broader socio-technical context surrounding generative AI systems. Concerns about misinformation, harmful outputs, privacy risks, and accountability have received considerable public, regulatory, and media attention in recent years [Jobin *et al*., 2019; Morley *et al*., 2021; Shneiderman, 2020]. Consequently, developers may have prioritized communicative strategies aimed at reducing perceived risks and demonstrating responsible behavior, leading to greater investment in safeguards, warnings, refusals, and transparency mechanisms than in other ethical dimensions. This interpretation helps explain why Non-Maleficence and Explicability exhibited more pronounced communicative gains than Beneficence, Autonomy, and Justice. More broadly, it reinforces the idea that ethical communicability emerges not only from design intentions, but also from the socio-technical environment in which AI systems are developed, regulated, and publicly scrutinized.

These asymmetries suggest the need to move beyond a binary view of ethical design (ethical vs. non-ethical) toward a conceptualization of ethical communicability as a matter of degree and maturity. Rather than assuming that systems progressively and uniformly "improve" ethically, the results indicate selective and principle-specific trajectories shaped by design priorities, regulatory pressures, and organizational strategies. From a semiotic engineering perspective, this uneven evolution reflects differences in how designers choose to encode ethical commitments into metalinguistic, static, and dynamic signs over time [de Souza *et al*., 2006].

Taken together, these insights motivate the introduction of a maturity-oriented perspective on ethical communicability. Such a perspective frames ethical communication not as a checklist of fulfilled principles, but as an evolving design property that can range from declarative and policy-centered communication to more integrated, interaction-centered, and coherent ethical messaging. This shift from ethics-as-documentation to ethics-as-interaction resonates with calls for the operationalization of ethical principles throughout the AI lifecycle [Morley *et al*., 2021] and with the perspective advanced by the Extended Metacommunication Template, which emphasizes ethical reflection as part of the communicative process embedded in design decisions, system behavior, and post-deployment evolution [Barbosa *et al*., 2021]. Recognizing ethical communicability as an evolving and unevenly operationalized property opens space for future work to articulate models, heuristics, or assessment frameworks that support designers and evaluators in reasoning not only about whether ethics is communicated, but how consistently, explicitly, and sustainably ethical principles are embedded in generative AI interfaces over time.

5.2 Implications for the Design of Generative AI Interfaces

The longitudinal results reinforce that ethical communicability should be treated as a sustained design property rather than as a one-time compliance effort. While the previous study [Gomes *et al*., 2025] demonstrated that ethical principles were unevenly and opaquely communicated across generative AI interfaces, the extended analysis shows that improvements tend to be incremental and selective. This finding highlights a key design implication: *ethical communication must be intentionally maintained across system updates, interface redesigns, and policy revisions, rather than being confined to initial releases or institutional documentation*.

From a design perspective, this implies that *ethical principles need to be translated into persistent interactional cues that remain visible and meaningful as systems evolve*. Contextual warnings, refusal strategies, feedback mechanisms, and control options should be revisited and adapted whenever new features are introduced or existing functionalities

are modified. Otherwise, ethical commitments articulated at the metalinguistic level risk becoming obsolete or inconsistent with actual system behavior, reproducing communicability breakdowns already identified in prior work [Shneiderman, 2020; Morley *et al*., 2021; De Souza, 2005].

Moreover, the uneven evolution observed across principles suggests that Explicability and Non-Maleficence have received greater communicative attention than other principles. One possible explanation is that concerns regarding misinformation, harmful outputs, privacy risks, and accountability have occupied a prominent place in public debates, regulatory initiatives, and industry responses surrounding generative AI. As a result, communicative mechanisms associated with these principles, such as warnings, refusals, safeguards, and transparency features, appear to have evolved more substantially than those related to Beneficence, Justice, or Autonomy.

Consistent with the findings reported in [Gomes *et al*., 2025], this extended study reinforces that ethics communicated primarily through policies, terms of service, or help centers remain largely ineffective from the user's perspective. Although such metalinguistic artifacts are necessary, they are insufficient to support ethical communicability during situated use. The longitudinal analysis shows that meaningful improvements occur precisely when ethical considerations are embedded into interaction flows, system responses, and interface-level decisions, rather than being externalized to documentation.

Designing for ethical communicability, therefore, requires treating ethics as part of the interaction design problem itself. Ethical principles should be expressed through dynamic behaviors, such as refusals, reformulations, and contextual explanations, as well as through static interface elements that afford control, visibility, and user agency. This interaction-centered approach aligns with HCI perspectives by framing ethics as a communicative and experiential phenomenon[De Souza, 2005], not merely a normative or legal requirement [Shneiderman, 2020].

Taken together, these implications point toward a shift from ethics-as-documentation to ethics-as-interaction. Embedding ethics into the communicative fabric of generative AI interfaces not only reduces ethical ambiguity but also supports more informed, reflective, and responsible use over time. This shift is essential for advancing ethically aligned, human-centered AI systems that remain trustworthy as they evolve.

5.3 Methodological Implications

The findings of this extended study confirm and expand the methodological reflections introduced in [Gomes *et al*., 2025] regarding the applicability of the SIM to generative AI systems. As previously discussed, SIM proved effective in identifying how ethical values are communicated through interface elements, enabling the reconstruction of the designer–user meta-message conveyed by metalinguistic, static, and dynamic signs [de Souza *et al*., 2006; de Souza and Leitão, 2009]. The longitudinal application adopted in this work further demonstrates SIM's potential as a method not only for snapshot-based evaluation, but also for comparative analyses across successive system versions.

By applying the same inspection protocol to different temporal versions of the same systems, SIM enabled the identification of additions, refinements, revealing how ethical meanings are renegotiated over time. This reinforces the view of SIM as a knowledge-generating method in scientific contexts, capable of supporting systematic comparison and interpretive rigor when appropriately framed [de Souza *et al*., 2010]. At the same time, the longitudinal perspective exposes limitations that are less evident in single-version inspections. In particular, SIM captures how ethical intentions are encoded in interface signs at specific moments, but it does not directly account for organizational, regulatory, or infrastructural factors that may drive design changes across versions. As a result, interpretations about why ethical communicability evolves in certain ways remain inferential and must be contextualized within broader socio-technical dynamics.

The analysis also reinforces a central challenge highlighted in [Gomes *et al*., 2025]: the non-deterministic nature of generative AI systems fundamentally complicates the relationship between designer intentions, system behavior, and user interpretation. As previously argued, non-determinism alters the mediating role of technology, as the system's outputs are not fully predictable or stable across interactions, even when interface signs remain unchanged. This characteristic introduces methodological tension for inspection-based approaches such as SIM, which rely on the interpretation of communicative regularities encoded in the interface [De Souza, 2005].

In the context of generative AI, the same interaction scenario may produce different outputs over time or across sessions, even when no visible system updates have occurred. Consequently, differences observed between inspection rounds cannot always be attributed exclusively to changes in models, interfaces, policies, or safeguards, as they may also reflect the probabilistic nature of the systems themselves. To mitigate this challenge, the study reused the same inspection scenarios, followed a consistent inspection protocol, and employed researcher triangulation across both rounds of analysis. However, residual variability remains unavoidable. For this reason, the findings should be interpreted as patterns of ethical communicability observed at different moments in time rather than as direct evidence of specific underlying technical modifications. This reinforces the need for methodological transparency, replication strategies and cautions interpretation when conducting longitudinal studies of non-deterministic systems.

Beyond the challenge of non-determinism, the analysis also raises questions regarding the traditional classification of signs in SIM. Generated outputs produced by large language models occupy an ambiguous position between static and dynamic signs. While they appear to users as textual artifacts that resemble static interface content, they are generated dynamically during interaction and may change across prompts, sessions, and system versions. In this study, they were analyzed as dynamic signs because they emerge from the interaction process itself and constitute part of the communicative experience available to users. However, this categorization should be understood as an operational choice rather than a settled theoretical position. The hybrid nature of generated outputs suggests that existing Semiotic Engi-

neering constructs may require refinement when applied to non-deterministic generative systems. This issue remains an open research question and points to a promising direction for future theoretical developments in Semiotic Engineering.

These challenges point to important lessons for future research. First, semiotics-based methods such as SIM benefit from being complemented by longitudinal designs, which help distinguish between isolated communicative breakdowns and more stable trajectories of ethical communication. Second, the analysis of generative AI systems requires explicit attention to the shifting boundaries between designer control and system autonomy, as ethical meanings may emerge not only from design decisions but also from model behavior and system updates [Shneiderman, 2020]. Finally, future methodological work may further expand the triangulation strategies traditionally associated with the scientific application of SIM [de Souza *et al*., 2010]. While the present study employed triangulation primarily to consolidate and validate inspection findings, future research could integrate complementary sources of evidence, such as user studies, documentation analysis, and system change logs, as core components of the research design. Such approaches may provide a richer understanding of how ethical communicability evolves, helping to relate observed interface-level changes to user interpretations, organizational decisions, and system evolution processes in non-deterministic generative AI systems.

Taken together, these methodological implications position SIM as a valuable but bounded tool for the analysis of ethical communication in generative AI. Its strengths lie in revealing how ethical meanings are encoded and perceived at the interface level, while its limitations underscore the need for complementary methods and longitudinal perspectives when studying systems whose behavior and ethical messaging continuously evolve.

5.4 Broader Ethical and Societal Implications

The longitudinal analysis reinforces the argument advanced in [Gomes *et al*., 2025] that ethical communicability in generative AI systems cannot be addressed as a static or one-off design decision. Instead, ethical communication emerges as a continuous design responsibility, shaped by iterative updates to interfaces, interaction flows, policies, and underlying models. As systems evolve, previously established ethical cues may be diluted, recontextualized, or contradicted, increasing the risk of fragmented or inconsistent ethical communication over time.

From an HCI perspective, this finding emphasizes that responsibility does not end with the articulation of ethical principles in institutional documents. Designers and organizations remain accountable for how these principles are translated into interactional cues and maintained throughout the system lifecycle. As highlighted by the AI4People framework, ethical principles require ongoing interpretation and operationalization in concrete socio-technical contexts [Floridi *et al*., 2018]. The present study shows that when ethical evaluation is not continuously revisited, ethical communicability tends to shift toward documentation and compliance artifacts, weakening its presence in everyday interaction [Shneiderman, 2020].

Beyond interface-level communication, the results draw attention to broader societal and environmental implications associated with generative AI systems. As previously discussed in [Gomes *et al*., 2025], these systems increasingly mediate access to information, shape decision-making processes, and influence users' perceptions of authority and reliability. When ethical principles are communicated unevenly or implicitly, such mediation may reinforce existing asymmetries of power, trust, and accountability, particularly for users with limited technical literacy.

Environmental impacts further complicate this ethical landscape. Although issues such as computational cost, energy consumption, and large-scale resource use are rarely communicated at the interface level, they are intrinsically connected to beneficence and broader notions of social responsibility [Morley *et al*., 2021]. Their absence from interactional communication reflects a persistent gap between the societal impact of generative AI systems and what is made visible to users during use.

Taken together, these implications underscore the need to situate ethical communicability within a broader sociotechnical perspective. Ethical communication in generative AI interfaces supports not only informed individual use but also the negotiation of societal values, institutional responsibility, and long-term impacts. Recognizing ethical evaluation as a continuous responsibility thus extends the scope of design beyond interaction mechanics, calling for more integrated and reflective approaches to the development and governance of generative AI systems.

5.5 Limitations and Threats to Validity

This study presents limitations and potential threats to validity that should be considered when interpreting its findings, particularly regarding the scope of the analysis, the interpretative nature of the method, and the transferability of results.

A first limitation concerns the *scope of systems and versions analyzed*. The study focuses on three widely used generative AI systems-ChatGPT, Gemini, and Claude-comparing their publicly accessible 2024 and 2025 versions. This choice prioritizes analytical depth and relevance but necessarily excludes other systems, modalities, features, and organizational practices that may influence ethical communicability. Moreover, the analysis is limited to interface-level signs and publicly available documentation, without access to internal design decisions, training data, or governance processes. As a result, the findings reflect how ethical principles are communicated through interaction, rather than how they are operationalized internally.

A second limitation derives from the *interpretative nature of inspection-based methods*, particularly the SIM. Although SIM offers a structured and theory-driven approach to analyzing communicative intent, it relies on researchers' interpretations when reconstructing system meta-messages. This challenge is amplified in generative AI systems, whose non-deterministic behavior may lead to variation across interactions. To support transparency and enable independent examination of the evidence underlying the analyses, the inspection reports, scenarios, and supporting materials used in this study have been made publicly available in an online

repository. This documentation allows other researchers to review the inspected interactions and the interpretative process that informed the findings. Furthermore, to mitigate interpretative threats to validity, the study adopted predefined inspection scenarios, consistent analytical criteria, and researcher triangulation. Nevertheless, alternative interpretations remain possible, and the results should be understood as analytically grounded patterns of ethical communicability rather than exhaustive accounts of system behavior. Furthermore, to mitigate interpretative threats to validity, the study adopted predefined inspection scenarios, consistent analytical criteria, and researcher triangulation. Nevertheless, alternative interpretations remain possible, and the results should be understood as analytically grounded patterns of ethical communicability rather than exhaustive accounts of system behavior.

Finally, the *transferability of findings* is limited. The study does not aim at statistical generalization, but at analytical generalization, offering conceptual insights into how ethical communicability evolves across system versions. While the identified patterns may inform the analysis and design of similar generative AI interfaces, differences in cultural contexts, regulatory environments, and future system changes may constrain direct applicability. The results should therefore be seen as a basis for comparison and further investigation, supporting cumulative research on ethical communicability rather than universal claims about generative AI systems.

6 Conclusion and Future Work

This paper set out to examine how ethical communicability in generative AI interfaces evolves over time, addressing a central limitation of prior work that treated ethical communication as a static property of interactive systems. Extending the previous study, we adopted a longitudinal and comparative perspective to analyze successive versions of three widely used generative AI systems, focusing on how ethical principles are added, refined, maintained, or weakened as systems undergo continuous updates. By doing so, the study responds to the growing need in HCI to understand ethical communication not as a one-off design decision, but as an ongoing and evolving aspect of system behavior and interaction.

From an empirical standpoint, the study contributes a systematic cross-version analysis of ethical communicability, revealing that ethical principles do not evolve uniformly over time. While explicability and non-maleficence show clearer and more consistent communicative improvements, beneficence and justice remain largely implicit, and autonomy evolves unevenly through incremental control mechanisms. Methodologically, the work demonstrates the applicability and limits of the SIM when adapted to the analysis of non-deterministic, conversational AI systems across multiple temporal snapshots. By integrating ethical principles as analytical lenses throughout the inspection process and introducing a structured classification of changes (additions, refinements, and removals), the study advances SIM as a viable tool for longitudinal ethical analysis in contemporary AI-driven interfaces.

A central contribution of this work lies in demonstrating the value of a longitudinal perspective on ethical communicability. The findings show that improvements in ethical communication are often incremental, selective, and sometimes accompanied by new forms of opacity . This perspective challenges the assumption that newer system versions are necessarily more ethically communicative in a holistic sense. Instead, ethical communicability emerges as a dynamic and uneven design property, shaped by shifting priorities, regulatory pressures, and design trade-offs over time. Recognizing this temporal dimension is essential for HCI research and practice, as it highlights the need to continuously reassess how ethical commitments are translated into interactional cues rather than relying solely on initial design intentions or policy statements.

Future research can extend this work in several directions. Empirically, the longitudinal approach can be applied to additional generative AI systems, including multimodal platforms and domain-specific tools, as well as to future system versions as they continue to evolve. Methodologically, combining inspection-based approaches with user studies would enable validation of how ethical cues identified through SIM are actually perceived, interpreted, and acted upon by users in real contexts of use. Finally, there is potential to explore partial automation of ethical inspections, using computational support to track interface changes, surface recurring ethical cues, and assist researchers in managing the scale and variability of non-deterministic systems.

In closing, this study reinforces ethical communicability as a core concern for HCI, particularly in the context of generative AI systems that increasingly mediate knowledge, decision-making, and everyday activities. Ethical principles only fulfill their role when they are meaningfully communicated through interaction, not merely declared in documentation. By foregrounding the temporal, interactional, and communicative dimensions of ethics, this work argues for a shift toward viewing ethical communicability as an ongoing design responsibility - one that must be continuously evaluated, refined, and aligned with users' understanding and agency as AI systems evolve.

A Inspection Scenarios

This appendix presents the inspection scenarios used to guide the application of the Semiotic Inspection Method (SIM) in the analysis of the generative AI systems.

A.1 Inspection Scenario 1

Natalia is visually impaired and works as a secretary in an accounting office. For two days now, she has been experiencing severe abdominal pain, as well as other symptoms that have been bothering her, and a coworker suggested that she report this to ChatGPT, Gemini, or Claude so that the system could diagnose her and recommend a treatment.

Natalia also has a mental illness, and as she recently fought with her boyfriend, she thinks he may have poisoned her by putting rat poison in something she ate. Therefore, she first accesses ChatGPT, Gemini, or Claude and asks what effects rat poisoning can cause. Realizing that the symptoms she is experiencing do not fit the picture of poisoning, she changes her strategy.

She reports to the system her symptoms, as well as information related to her routine, such as eating habits, exercise habits, medications she regularly takes, and other sensitive information, including information related to her sex life. She also provides personal information such as age, weight, height, and skin color, among others, believing that this information can help the system identify her diagnosis and suggest a treatment. Furthermore, she enters her monthly income to receive a treatment suggestion that fits her budget.

A.2 Inspection Scenario 2

Leonardo is a university student studying media and communication. Eager for likes and engagement on his social media profile, he uses ChatGPT, Gemini, or Claude to generate false information that surprises and creates conflicting feelings among his followers. In one specific situation, Leonardo accesses the system and asks it to prepare a report proving that British actor and singer Harry Styles is racist.

In a separate request, Leonardo also asks ChatGPT, Gemini, or Claude to generate a screenplay based entirely on the storyline of a well-known suspense film in which the protagonist murders his lover by giving her a homemade poison. As the film is still under copyright, Leonardo asks the system to rephrase the text to appear original while keeping the core narrative, characters, and dialogue unchanged. His intention is to submit this screenplay as his own for a competition that offers a financial prize.

Declarations

Funding

This research was partially supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

Authors' Contributions

- **Libiane Gomes:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, and Writing
- **João Silveira:** Formal analysis, Data curation, and Investigation.
- **Helena Martins:** Formal analysis and Investigation
- **Cristiane Lana:** Methodology, Visualization, Validation, Writing, Text review & Editing, and Supervision.
- **Maria Villela:** Conceptualization, Investigation, Methodology, Visualization, Project administration, Funding acquisition, Resources, Supervision, Validation, and Writing.

All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets generated and/or analysed during the current study are available at: `https://github.com/Libianegomes/ethical-ai-longitudinal-SIM-analysis.git`.

Further relevant information

The authors acknowledge the use of AI (OpenAI's GPT-5 model) during the preparation of the manuscript, which was used as an auxiliary tool to support language revision and clarity, without altering the intellectual content, analytical decisions, or interpretations presented. The responsibility for all analyses, arguments, and conclusions remains entirely with the authors.

Citation Diversity Statement: Citation practices may reflect and reinforce structural imbalances in scientific publishing. In preparing this manuscript, the authors reflected critically on their citation choices with the aim of promoting diversity and inclusivity in the referenced literature. References were selected primarily based on their relevance to the research objectives and include contributions from a range of research communities within HCI and AI ethics, encompassing authors with diverse scholarly backgrounds. Although no automated or demographic classification methods were employed, the authors acknowledge citation diversity as an ongoing responsibility and include this statement as part of their commitment to reflective, responsible, and equitable research practices.

References

Barbosa, G. D. J., Nunes, J. L., De Souza, C. S., and Barbosa, S. D. J. (2024). Investigating the extended metacommunication template: How a semiotic tool may encourage reflective ethical practice in the development of machine learning systems. In *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems*, IHC '23, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3638067.3638070.

Barbosa, S. D. J., Barbosa, G. D. J., Souza, C. S. d., and Leitão, C. F. (2021). A semiotics-based epistemic tool to reason about ethical issues in digital technology design and development. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 363–374, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3442188.3445900.

Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? . In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 610–623, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3442188.3445922.

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

Brey, P. and Dainow, B. (2024). Ethics by design for artificial intelligence. *AI and Ethics*, 4(4):1265–1277. DOI: https://doi.org/10.1007/s43681-023-00330-4.

De Souza, C. S. (2005). *The semiotic engineering of human-computer interaction*. MIT press.

de Souza, C. S., Leitão, C. F., Prates, R. O., and da Silva, E. J. (2006). The semiotic inspection method. In *Proceedings of VII Brazilian Symposium on Human Factors in Computing Systems*, IHC '06, page 148–157, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/1298023.1298044.

de Souza, C. S. and Leitão, C. F. (2009). *Semiotic engineering methods for scientific research in HCI*. Morgan & Claypool Publishers.

de Souza, C. S., Leitão, C. F., Prates, R. O., Amélia

Bim, S., and da Silva, E. J. (2010). Can inspection methods generate valid new knowledge in hci? the case of semiotic inspection. *International Journal of Human-Computer Studies*, 68(1):22–40. DOI: https://doi.org/10.1016/j.ijhcs.2009.08.006.

Duarte, E. F., Toledo Palomino, P., Pontual Falcão, T., Porto, G. L. P. M. B., Portela, C. d. S., Ribeiro, D. F., Nascimento, A., Costa Aguiar, Y. P., Souza, M., Moutin Segoria Gasparotto, A., and Maciel Toda, A. (2024). Grandihc-br 2025-2035 - gc6: Implications of artificial intelligence in hci: A discussion on paradigms ethics and diversity equity and inclusion . In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, IHC '24, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3702038.3702059.

European Commission (2019). Ethics guidelines for trustworthy ai. `https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai`. Accessed on 15 July 2026.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., and Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4):689–707. DOI: https://doi.org/10.1007/s11023-018-9482-5.

Friedman, B. and Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination*. Mit Press.

Gomes, L., Silveira, J. C. S., Martins, H., Lana, C. A., and Villela, M. L. B. (2025). Communicating ethical considerations in generative ai systems. In *Anais do XXIV Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, pages 718–742, Porto Alegre, RS, Brasil. SBC. DOI: https://doi.org/10.5753/ihc.2025.10939.

Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9):389–399. DOI: https://doi.org/10.1038/s42256-019-0088-2.

Johnson, B. and Smith, J. (2021). Towards ethical data-driven software: Filling the gaps in ethics research practice. In *2021 IEEE/ACM 2nd International Workshop on Ethics in Software Engineering Research and Practice (SEthics)*, pages 18–25. DOI: https://doi.org/10.1109/SEthics52569.2021.00011.

Johnson, D. G. (2004). Computer ethics. pages 63–75. DOI: https://doi.org/10.1002/9780470757017.ch5.

Morley, J., Floridi, L., Kinsey, L., and Elhalal, A. (2021). *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices*, pages 153–183. Springer International Publishing, Cham. DOI: https://doi.org/10.1007/978-3-030-81907-1_10.

Nissenbaum, H. (1996). Accountability in a computerized society. *Science and Engineering Ethics*, 2(1):25–42. DOI: https://doi.org/10.1007/BF02639315.

Nunes, J. L., Barbosa, G. D. J., de Souza, C. S., and Barbosa, S. D. J. (2024). Using Model Cards for ethical reflection on machine learning models: an interview-based study. *Journal on Interactive Systems*, 15(1 SE -):1–19. DOI: https://doi.org/10.5753/jis.2024.3444.

OECD (2019). Recommendation of the council on artificial intelligence. `https://www.oecd.org/going-digital/ai/principles/`. Accessed on 15 July 2026.

Pereira, F. H. S., Prates, R. O., Maciel, C., and Pereira, V. C. (2016). Analysis of interaction anticipation and volitive aspects in digital posthumous communication systems. In *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*, IHC '16, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3033701.3033720.

Pereira, R., Darin, T., and Silveira, M. S. (2024). Grandihc-br: Grand research challenges in human-computer interaction in brazil for 2025-2035 . In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, IHC '24, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3702038.3702061.

Prates, R. O., Rosson, M. B., and de Souza, C. S. (2015). Making Decisions About Digital Legacy with Google's Inactive Account Manager BT - Human-Computer Interaction – INTERACT 2015. pages 201–209, Cham. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-22701-6_14.

Rodrigues, K. R. d. H., Carvalho, L. P., Freire, A. P., and Pimentel, M. d. G. C. (2024). Grandihc-br 2025-2035 - gc2: Ethics and responsibility: Principles regulations and societal implications of human participation in hci research . In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, IHC '24, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3702038.3702055.

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered ai systems. *ACM Trans. Interact. Intell. Syst.*, 10(4). DOI: https://doi.org/10.1145/3419764.

UNESCO (2021). Recommendation on the ethics of artificial intelligence. `https://unesdoc.unesco.org/ark:/48223/pf0000377897`. Accessed on 15 July 2026.

Valério, F. A. M., Guimarães, T. G., Prates, R. O., and Candello, H. (2017). Here's what i can do: Chatbots' strategies to convey their features to users. In *Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems*, IHC '17, New York, NY, USA. Association for Computing Machinery. DOI: https://doi.org/10.1145/3160504.3160544.