

Article

Not peer-reviewed version

The Information Efficiency Metric (IEM): An Info-Metric Approach for Quantifying AI Language Model Performance

[Ljerka Luić](#), [Maja Barbić](#), [Marijana Rončević](#) *

Posted Date: 25 February 2026

doi: 10.20944/preprints202602.1562.v1

Keywords: info-metrics; information theory; digital intelligence; human-AI communication; large language models; information efficiency; quantitative analysis; conversational AI; entropy-based metrics



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Article

The Information Efficiency Metric (IEM): An Info-Metric Approach for Quantifying AI Language Model Performance

Ljerka Luić, Maja Barbić and Marijana Rončević *

University North, University Center Koprivnica, Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croatia

* Correspondence: maroncevic@unin.hr

Abstract

The interaction between humans and artificial intelligence has become a critical channel for information exchange, yet no quantitative, theoretically grounded framework exists for measuring information efficiency in human–AI communication. This study empirically validated an info-metrics framework operationalizing information efficiency through three dimensions: information density (D), relevance (R) and redundancy (Q) synthesized into an information efficiency metric (IEM). We analyzed 60 AI responses from ChatGPT 5.2 and Claude Opus 4.5 across factual, analytical and creative question types using combined coding, automated structural measures and human evaluation of informational units. Results showed that information density and relevance positively contributed to IEM, while redundancy had a negative contribution. Efficiency varied by task type, with factual prompts showing highest variability across models and highest efficiency. Contrary to expectations, creative responses did not exhibit higher redundancy suggesting that expressive diversity does not necessarily constitute informational noise. The framework offers a task-sensitive, theoretically grounded approach to evaluate human–AI information exchange beyond correctness or subjective quality judgement, supporting systems-oriented optimization of conversational AI protocols.

Keywords: info-metrics; information theory; digital intelligence; human–AI communication; large language models; information efficiency; quantitative analysis; conversational AI; entropy-based metrics

1. Introduction

The rapid development of artificial intelligence (AI) and large language models (LLMs) has fundamentally changed the way we communicate and access information. Human–AI communication today takes place through interfaces that mimic human communication, making information and knowledge more accessible than ever before. These technological changes create a need to understand and value this new form of information exchange. In modern digital environments, communication between humans and artificial intelligence can be understood as a complex information process in which AI systems participate not only as technical intermediaries, but also as active participants in communication [1]. Users often interpret AI responses as meaningful and purposeful communication, although AI systems do not possess human-like intent, which further affects information perception and decision-making [2]. Recent theoretical discussions emphasize the need to develop communication theories for AI as a dialogue partner that go beyond traditional transmission models [3].

Conversational AI systems are increasingly becoming communication participants and mediators within the information search and decision-making processes in education and creative production [4]. Human–AI interaction is no longer a purely technical matter of choosing communication tools. It has become a structured communication and information process with direct

social and organizational consequences, exerting a growing influence on overall society [5,6]. Such a transformation suggests that AI-mediated communication is becoming an interdisciplinary field that connects information theory, communication science and research on human-machine interaction [7]. The application of information theory has evolved from Shannon's original formulations for technical communication channels to the increasingly complex domains of digital media and, more recently, human-AI interaction [8]. Relational elements such as trust, anthropomorphization, and perceived competence of AI systems further shapes the way users accept information [9]. However, the approach to understanding and evaluating human-AI communication is the subject of significant theoretical and methodological debate. Traditionally, digital intelligence has been presented in literature as a set of skills required to operate in digitally mediated environments. A contemporary perspective proposes that digital intelligence be viewed as an emergent property of human-AI interaction itself, where the quality of the outcome depends not only on the user's skills, but also on the way the AI structures, filters, and generates information. This conceptual distinction has far-reaching implications for the way we assess the effectiveness of AI systems. In addition, the application of classical information theory [10] to semantically rich communication is the subject of academic debate. While the traditional perspective maintains a focus on the statistical properties of signals, contemporary approaches emphasize the necessity of extending it to the semantic and pragmatic dimensions, where information value depends on context, relevance, and user intent [11].

Recent research highlights the potential for collaborative development that combines the benefits of AI with human domain knowledge to improve desired outcomes and create value [12]. This approach is particularly relevant in creative work, where AI acts as an active participant and partner. Successful symbiosis requires understanding the quality of information exchange that takes place within the partnership, especially in the context of objectively evaluating the quality of AI contributions [13]. Empirical research on human-AI collaboration further shows that the effectiveness of interaction depends not only on the technical characteristics of the AI system, but also on the protocols through which the joint work takes place. Cabitza et al. [14] confirmed that aggregated human judgments can outperform the results of AI systems, suggesting that the quality of outcomes stems from the structure of the interaction itself. A key question arising from the symbiotic perspective is how to measure the success of such a partnership. Research shows that the correctness of outcomes is often determined by the underlying technology and the way in which performance and user satisfaction are shaped by the collaborative nature of the system [12], which suggests the need for objective measures that can quantify the contribution of AI partners. Generative AI systems are becoming increasingly independent in creating, selecting, and structuring information, making it necessary to view digital intelligence in terms the efficiency of information exchange in human-AI communication. The key research question is no longer simply whether AI systems produce correct answers, rather how informationally efficient these systems are in conveying meaningful content, to what extent they remain relevant to user queries, and to what level of redundancy they introduce into communication. Such a shift in focus requires a theoretical and methodological framework that allows for the evaluation of the information structure of AI responses, i.e., their effectiveness in conveying meaningful content with minimal redundancy, rather than simply assessing the semantic or functional quality of the output [15]. Information theory provides a fundamental framework for understanding uncertainty, entropy, and the efficiency of communication channels [10]. Shannon's model, originally developed for technical communication systems, defines the key concepts: a source sends a message through a channel that may contain noise, to a receiver. Entropy measures the average uncertainty of the source of the message and represents the amount of information needed to accurately encode the message. Channel capacity denotes the maximum amount of information that can be reliably transmitted, and communication efficiency is implicitly viewed as the ability of a system to transmit as much information as possible with minimal loss [16]. However, when applied to human-AI communication, this approach shows limitations because it ignores the semantic and pragmatic dimensions of communication [11]. Modern interpretations of information theory emphasize the need to extend the classical framework

to the semantic and pragmatic levels [17], where information is viewed as meaningful content that reduces uncertainty in relation to a specific goal.

A key concept emerging from Shannon's theory is the relationship between signal and noise. In the context of human–AI communication, noise does not necessarily appear as a technical nuisance, but as semantic and pragmatic excess: redundant formulations, repetitions, irrelevant digressions. Such an interpretation allows for a quantitative observation of the effectiveness of AI responses not only in terms of accuracy, but also in terms of their ability to economically and precisely convey meaningful content, thereby aligning the evaluation of AI systems with theoretical approaches that emphasize the role of meaning, context, and purpose in AI-generated communication [15]. Info-metrics [18] extends Shannon's framework (1948) [10] by introducing constraints, context, and decision goals. Developed through the work of Amos Golan and colleagues, it is based on the premise that "info-metrics" is the science of modeling, reasoning, and deriving information under conditions of uncertainty and limited data [19]. Unlike classical statistics, which relies on assumptions about distributions and large samples, info-metrics starts from the principle of entropy maximization under given constraints by choosing the "most neutral" possible distribution. This makes it possible to draw conclusions even when the data is incomplete, noisy or partially known. The central principle of info-metrics is the maximization of entropy while respecting the available information in the form of constraints. These constraints can be empirical (e.g., observations) or conceptual (e.g., logical relationships, goals). This approach has proven successful in various fields, including the quantification of information dissemination in social networks where Shannon entropy is applied to different dimensions of the information process: time interval, users, content and location, to quantify the dynamics of information dissemination [20]. Analogously, in this paper, we apply entropy principles to the structure of AI-generated responses, where information density, relevance, and redundancy represent constraints within which information efficiency is maximized.

In the context of AI communication, the user query, the task context, and the expected outcome can be interpreted precisely as a set of constraints within which the AI generates a response. This concept is analogous to the approach in the field of human–AI collaborative optimization, where human expert knowledge is operationalized as a set of constraints that guide the process towards the optimal solution [21]. Just as expert knowledge can guide AI in the optimization process, the user query defines the relevant information space within which the AI system generates a response. Info-metrics proves to be particularly suitable for the analysis of human–AI interactions for several reasons: AI responses are probabilistic, not deterministic, user queries explicitly define what is relevant, naturally model the trade-off between informativeness and redundancy, and do not require a predefined "correct" answer, which is crucial for open-ended and creative tasks. This perspective is consistent with research in the field of artificial intelligence security, which emphasizes the need to evaluate AI systems in complex and poorly specified environments without relying on strictly defined target outcomes [22]. Info-metrics thus allows for the quantitative evaluation of information efficiency without reducing the complexity of the communication context. However, these formal approaches have not been systematically applied to the assessment of information efficiency in human–AI communication. Shannon's theory applied to conversational AI systems shows significant limitations: it assumes a one-way communication model while human–AI communication is iterative and contextual, ignores meaning, does not include the goal of communication, and treats redundancy exclusively as noise. This gap becomes particularly relevant because AI systems today act as active information brokers that structure and present information, thus shifting the focus of digital intelligence from individual competencies to the quality of information exchange within the human–AI interaction itself. Despite the development of numerous evaluation metrics for assessing human–AI interaction, existing approaches mainly rely on response accuracy, language quality or user satisfaction [23–25], while explicitly modeling the relationship between information density, relevance and redundancy is neglected. This lacks an information-theoretic perspective based on the signal-to-noise ratio [10]. The importance of developing objective measures of information efficiency is further confirmed by recent research showing that AI systems can simultaneously reduce human

cognitive load and mitigate user heuristic biases, as well as increase confirmation bias when AI recommendations match erroneous human judgments [26]. Rosbach et al. [27] empirically confirmed the phenomenon of false confirmation in the context of computer pathology. These findings point to the need for evaluation frameworks that allow assessing the information quality of AI responses.

The main objective of this paper is to develop and empirically validate an info-metric model for measuring information efficiency in human–AI communication, operationalized through the information efficiency metric (IEM) that integrates information density, relevance, and redundancy. In accordance with the stated objective, the following hypotheses are formulated:

- **H1:** Information density (D) and relevance (R) have a positive effect on information efficiency (IEM), while redundancy (Q) has a negative effect on information efficiency (IEM).
- **H2:** The variability in information efficiency (IEM) between AI models is greater for factual questions than for analytical and creative questions.
- **H3:** Responses to creative questions will have a statistically significantly higher level of redundancy (Q) compared to responses to factual and analytical questions.
- **H4:** Responses to factual questions achieve higher information efficiency (IEM) compared to responses to analytical and creative questions.

The paper introduces an information perspective on digital intelligence, conceptualizing it as an emergent property of human–AI interaction operationalized through measurable information exchange efficiency. The results confirm that information density, relevance, and redundancy systematically influence the overall information efficiency of AI responses.

2. Literature Review

Classical information theory and info-metrics appeared in scientific literature in clearly separable historical phases, related to different epistemological and methodological needs of science. The term information theory appears for the first time explicitly and systematically in the literature in the work of C. Shannon [10], who defined a quantitative approach to understanding the transmission of messages through technical channels. In contrast, info-metrics appear much later, and evolve in the context of applied statistics and economics, focusing on inference with limited information, in which A. Golan [28] played a key role. The contemporary application of info-metric principles to the evaluation of digital intelligence represents a natural evolution, expanding the reach from the technical parameters of transmission to the semantic-pragmatic dimension of communication. However, although info-metrics represent a methodological framework that uses Shannon's entropy and information measures, they have not yet appeared in the literature in terms of measuring the information efficiency of human–AI communication. To position the proposed information efficiency framework within a broader research area, a systematic literature review (SLR) method was conducted according to the PRISMA 2020 guidelines [29]. The aim of the review was to identify relevant research dealing with info-metric approaches, information efficiency measures and the evaluation of human–AI communication. The search was conducted in two major scientific databases: Scopus and Web of Science (WoS), where the title, abstract and keyword fields were searched to ensure a high level of thematic relevance. The search strategy was based on four thematic units that encompass the full range of relevant literature:

1. Information-theoretical approaches to human–AI communication
2. Info-metric approaches in the AI context
3. Digital intelligence and measurement of efficiency
4. Communication efficiency of conversational systems

Detailed search strategies for both databases are shown in Tables 1 and 2.

Table 1. Scopus search strategy and results.

Search strategy	Hits	Timeframe	Index
TITLE-ABS-KEY((((“information theory” OR “entropy” OR “information efficiency” OR “information density” OR “redundancy”) AND (“human-AI interaction” OR “human-computer interaction” OR “digital intelligence” OR “AI communication”) AND (“evaluation” OR “metric” OR “quantitative analysis” OR “performance measurement”)) OR (“info-metrics” OR “maximum entropy econometrics” OR “entropy-based inference”) AND (“artificial intelligence” OR “decision-making” OR “data modeling”)) OR (“digital intelligence” OR “AI quotient” OR “AI literacy” OR “human-AI collaboration”) AND (“measurement” OR “quantification” OR “index” OR “metric” OR “efficiency”)) OR (“communication efficiency” OR “information transfer” OR “semantic efficiency” OR “information redundancy”) AND (“AI” OR “language model” OR “conversational agent” OR “chatbot”))) AND PUBYEAR > 2014	1595	All years	Scopus
Refined by: (TITLE-ABS-KEY (((“information theory” OR “entropy” OR “information efficiency” OR “information density” OR “redundancy”) AND (“human-AI interaction” OR “human-computer interaction” OR “digital intelligence” OR “AI communication”) AND (“evaluation” OR “metric” OR “quantitative analysis” OR “performance measurement”)) OR ((“info-metrics” OR “maximum entropy econometrics” OR “entropy-based inference”) AND (“artificial intelligence” OR “decision-making” OR “data modeling”)) OR ((“digital intelligence” OR “AI quotient” OR “AI literacy” OR “human-AI collaboration”) AND (“measurement” OR “quantification” OR “index” OR “metric” OR “efficiency”)) OR ((“communication efficiency” OR “information transfer” OR “semantic efficiency” OR “information redundancy”) AND (“AI” OR “language model” OR “conversational agent” OR “chatbot”))) AND PUBYEAR > 2014) AND PUBYEAR > 2021 AND PUBYEAR < 2027 AND (LIMIT-TO (DOCTYPE , “ar”) OR LIMIT-TO (DOCTYPE , “bk”)) AND (LIMIT-TO (LANGUAGE , “English”)) AND (LIMIT-TO (SUBJAREA , “COMP”) OR LIMIT-TO (SUBJAREA , “SOCI”) OR LIMIT-TO (SUBJAREA , “DECI”) OR LIMIT-TO (SUBJAREA , “MULT”) OR LIMIT-TO (SUBJAREA , “ENGI”) OR LIMIT-TO (SUBJAREA , “BUSI”))	530	2020-2026	Scopus

Table 2. WOS search strategy and results.

Search strategy	Hits	Timeframe	Index
((“information theory” OR “entropy” OR “information efficiency” OR “information density” OR “redundancy”) AND (“human-AI interaction” OR “human-computer interaction” OR “digital intelligence” OR “AI communication”) AND (“evaluation” OR “metric” OR “quantitative analysis” OR “performance Measurement”)) OR ((“info-metrics” OR “maximum entropy econometrics” OR “entropy-based inference”) AND (“artificial intelligence” OR “decision-making” OR “data modeling”)) OR ((“digital intelligence” OR “AI quotient” OR “AI literacy” OR “human-AI collaboration”) AND (“measurement” OR “quantification” OR “index” OR “metric” OR “efficiency”)) OR ((“communication efficiency” OR “information transfer” OR “semantic efficiency” OR “information redundancy”) AND TITLE-ABS-KEY (“AI” OR “language model” OR “conversational agent” OR “chatbot”))	466	All years	CPCI-S, SCI-EXPANDED, ESCI, CPCI-SSH, SSCI, BKCI-S

Refined search: DOCUMENT TYPES: (ARTICLE OR PROCEEDINGS PAPER) AND LANGUAGES: (ENGLISH) AND PUBLICATION YEARS: (2026 OR 2025 OR 2024 OR 2023 OR 2022 OR 2021 OR 2020) AND WEB OF SCIENCE CATEGORIES: (COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR COMPUTER SCIENCE INFORMATION SYSTEMS OR COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS OR COMMUNICATION OR COMPUTER SCIENCE THEORY METHODS OR INFORMATION SCIENCE LIBRARY SCIENCE OR MULTIDISCIPLINARY SCIENCES) AND RESEARCH AREAS: (COMPUTER SCIENCE OR COMMUNICATION OR BUSINESS ECONOMICS OR INFORMATION SCIENCE LIBRARY SCIENCE OR SCIENCE TECHNOLOGY OTHER TOPICS OR SOCIAL SCIENCES OTHER TOPICS)

CPCI-S, SCI-EXPANDED, ESCI, CPCI-SSH, SSCI, BKCI-S

154 2020-2026

Inclusion criteria:

- publication year: 2020-2026
- type of publication: scientific articles and book chapters
- language: English
- research areas: computer science, social sciences, information science, communication, engineering, business

Exclusion criteria:

- articles in the fields of medicine, chemistry, neuroscience, and physics (unless directly relevant to AI communication)
- publications without a peer-review process
- articles that do not address human–AI communication or evaluation of AI systems
- conference abstracts, editorials, and reviews without original research

The initial database search was conducted in December 2025 in two scientific databases: Scopus and Web of Science, and resulted in 1595 records from Scopus and 466 records from Web of Science (WoS), a total of 2061 records. In the title and abstract screening phase, papers that were not thematically relevant to info-metric approaches or human–AI communication were excluded. This screening resulted in 420 records from Scopus and 154 records from WoS that passed to the stage of detailed reading of the full text. In the final phase of the full-text analysis, papers were evaluated according to the following criteria:

- the presence of an explicit info-metric or information-theoretic framework
- focus on quantification or measurement of information efficiency
- relevance to human–AI communication or evaluation of AI systems
- empirical validation of proposed measures (where applicable)

After a detailed review and removal of studies that did not meet the final criteria, and after adding relevant literature from the reference lists of the selected papers, the final 54 records were included in the qualitative synthesis.

Figure 1 shows the PRISMA flow chart of a systematic literature review [29].

Despite the large number of papers in the broader field of AI system evaluation and human–AI communication (684 papers after applying basic filters), a systematic literature review showed that there is no quantitative, theoretically grounded info-metric model for measuring the information efficiency of human–AI communication.

A qualitative synthesis of the identified literature allowed for the categorization of approaches into five main groups:

1. Evaluation of AI systems based on accuracy and performance (~40% of papers): Papers that focus on technical metrics such as precision, responsiveness, F1-score, and accuracy, but do not address the information structure of the response or the efficiency of information transfer.
2. Subjective user experience measures (~25% of papers): Research that uses questionnaires and scales to assess user satisfaction, perceived usefulness, and trust in AI systems, but without a quantitative information framework.
3. Semantic and linguistic metrics (~20% of papers): Approaches based on measuring similarity to reference texts, semantic vectors, or response coherence, without explicitly modeling information density, relevance, and redundancy.
4. Theoretical papers on digital intelligence (~10% of papers): Conceptual discussions on digital literacy, AI competencies, and human–AI collaboration, without operationalization through quantitative info-metric measures.
5. Other approaches (~5% of papers): Research focused on trust in AI, ethical aspects, explainability, cognitive load, and biases of AI systems.

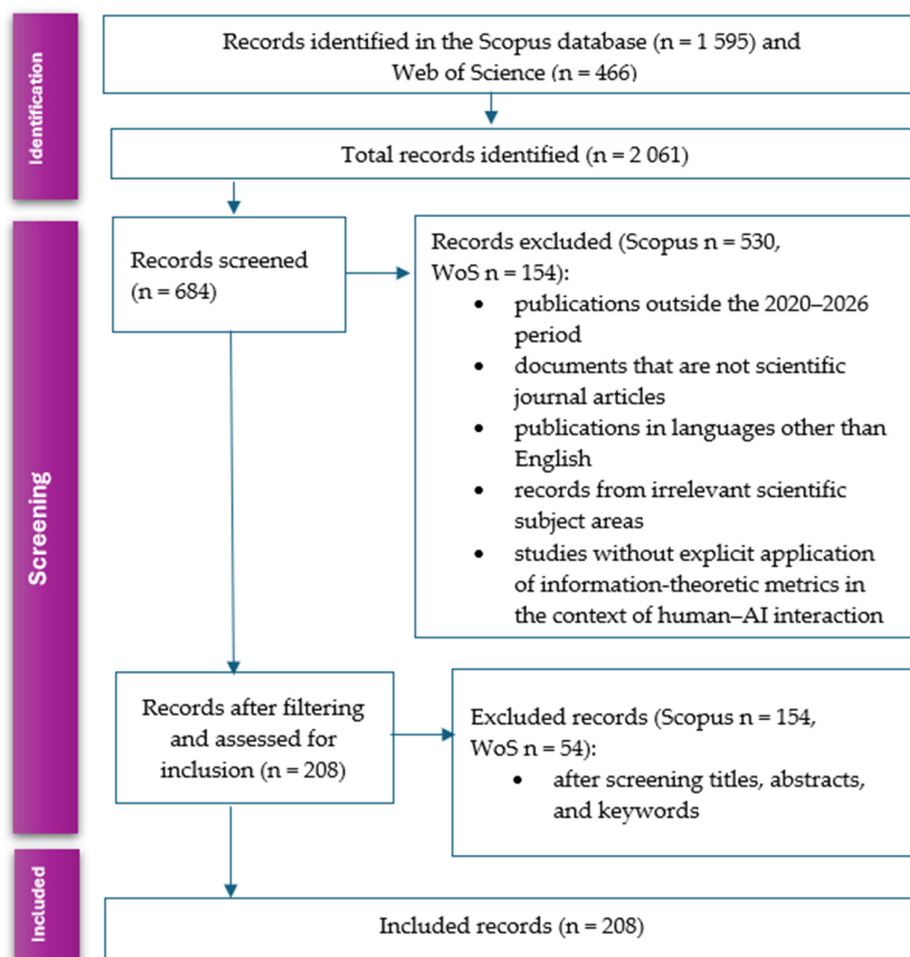


Figure 1. PRISMA flow chart of the systematic literature review.

A key finding of the systematic review is the identification of a clear research gap: while there are numerous approaches to evaluating AI systems and measuring user experience, none apply a systematic info-metric framework to quantify information efficiency in human–AI communication. Existing approaches focus on output accuracy, subjective satisfaction, or semantic similarity, neglecting fundamental information aspects such as signal-to-noise ratio, information density, and

redundancy in AI responses. This gap motivates the development of the info-metric model proposed in this paper.

3. Conceptual Model Design

3.1. Conceptual Framework

Human–AI communication can be conceptualized as an information process that begins with a user query, where the query acts as an initial signal and a framework of constraints within the interaction. The AI system then generates a response as an information outcome shaped by uncertainty and probabilistic reasoning, producing a combination of informative and redundant segments due to limited insight into the user’s intent and the absence of real-time truth-checking.

The proposed framework evaluates generated content through three dimensions of communication effectiveness: the information structure of the response, its relevance to the user’s query, and the level of content that does not contribute to the communication goal. Their synthesis into the information efficiency metric (IEM) allows different AI models and query types to be compared on a common basis, revealing patterns overlooked in evaluations focused solely on accuracy or style. Conceptually, this framework connects the classical information-theoretic idea of the efficiency of information transmission with an info-metric approach to constraints and outcomes, applying them to human–AI interaction as a measurable information process within the contemporary infosphere [5].

3.2. Conceptual Model Design

The starting point of the proposed model is the assumption that human–AI communication can be analyzed as an information process whose quality is not limited to the accuracy and linguistic and stylistic acceptability of the response. The key research challenge is therefore not only the content of AI responses, but their information efficiency in transmitting meaningful content, which is in line with recent research on generative AI, which emphasizes the need to evaluate AI systems through the quality of generated content and its alignment with user goals [30].

Since existing evaluation methods rarely operationalize the signal-to-noise ratio, and information theory does not offer direct measures for semantically rich dialogic systems, this paper develops an info-metric model that measures communication efficiency through the structure of the generated response, rather than through its surface quality.

The model is based on three complementary fundamental dimensions of information exchange: information density (D) as the information value of the content, relevance (R) as the correspondence of the response to the query, and redundancy (Q) as the inefficiency caused by unnecessary repetition, which are integrated into a single information efficiency metric (IEM) as a measurable and comparable framework for assessing the information efficiency of AI responses.

3.2.1. Information Density

Information density (D) is defined as the ratio of semantically unique information elements to the total length of the response, operationalizing the assumption that a larger amount of text does not necessarily imply a larger amount of information. In the context of human–AI communication, D reflects the ability of the system to generate information-efficient responses with minimal content redundancy.

Formally, information density can be expressed as:

$$D = \frac{I_u}{L} \quad (1)$$

where I_u is the number of unique semantic units and L is the response length (number of words within the response). A high value of information density (D) indicates information-efficient responses, while low values signal content dilution.

3.2.2. Relevance

Relevance (R) indicates the level of correspondence between the response and the user query, and measures how well the AI maintains focus on the communication goal. It can be operationalized by a combination of human assessment and automated semantic measures, with normalization in the interval [0,1] where a higher value indicates a higher correspondence (relevance) of the response to the query.

Formally, relevance can be expressed as:

$$R = \frac{I_{rel}}{I_u} \quad (2)$$

where I_{rel} is the number of relevant semantic units compared to the number of unique semantic units. Theoretically, relevance represents a constraint in an info-metric sense. The user query acts as a constraint that defines the space of permissible information, while relevance measures the retention of the response within that space. This understanding has parallels in research on human–AI collaborative optimization, where explicit constraints narrow the space of possible solutions and enable more efficient convergence towards the optimal outcome [21].

3.2.3. Redundancy

Redundancy (Q) indicates the proportion of repetitive or low-informative segments within the response. In classical theory of information redundancy is viewed as noise [10], while in this model it is allowed that limited redundancy can have an explanatory function.

Redundancy is formally defined as:

$$Q = \frac{I_r}{L} \quad (3)$$

where I_r represents the number of redundant units and L is the length of the response. Higher values of Q indicate higher information noise and lower efficiency.

3.2.4. Definition of Information Efficiency Metric (IEM)

In Shannon's theory of information, information efficiency is traditionally associated with the signal-to-noise ratio within a communication channel [10]. However, in human–AI communication, the channel is not only a technical medium of transmission but also includes the communication outcome in the form of a generated textual response. Therefore, in this environment, information efficiency cannot be assessed solely through the accuracy of the response or the writing style, but requires an analysis of the response structure and the relationship between the cost-effectiveness and usefulness of the information produced by the AI system.

From an info-metrics perspective, a user query defines a set of constraints that determine the space of relevant information, while the AI response represents the final information outcome within the framework of that space under conditions of uncertainty and probabilistic reasoning [18]. An effective response must therefore simultaneously contain sufficient information content, remain consistent with the query, and minimize noise in the form of redundancy.

On this basis, measuring the efficiency of human–AI communication through information efficiency metrics (IEM) should take into account the relationship between the amount of information produced, how well this information complies with given constraints, and the level of information noise, i.e., connect three key dimensions of information exchange: information density (D), relevance (R), and redundancy (Q).

The following assumptions underline the structure of the metric:

1. Information efficiency assumes economy and utility. A larger amount of text does not mean greater informativeness.
2. Information brings value only if it is directed towards the goal, that is, if it is relevant to the user's query.

3. Excess content that does not contribute to the goal acts as information noise, increasing the communication cost and cognitive load of the user.

In accordance with the assumptions, the information efficiency metric (IEM) is defined as:

$$IEM = \frac{D \times R}{(1 + Q)} \quad (4)$$

The multiplicative relationship between information density and relevance ensures that informativeness is valued only if it is also aligned with the user's query. In this way, the idea is formalized that an information "signal" in human-AI communication is created only when the content is both dense and relevant. Redundancy appears in the metric as a penalizing factor, as it increases the noise level and reduces the overall efficiency of communication.

The three components of the IEM metric can be analytically linked to key concepts of Shannon's information theory [10]. Information density (D) conceptually corresponds to the entropy of the message source, as it describes the amount of potentially useful information contained in the generated response relative to its length. Relevance (R) can be interpreted as an analogue of mutual information, i.e., the measure to which the generated content truly corresponds to the user's query and reduces uncertainty in relation to the given communication goal. Redundancy (Q) here takes on the role of information noise, as it denotes the proportion of content that does not contribute to the achievement of communication purpose. Such a mapping does not imply a direct mathematical equivalence with the logarithmic structures of Shannon's entropy, but rather a functional analogy. The IEM metric retains the central information-theoretic principle of the signal-to-noise ratio, operationalizing it in the form of a normalized and empirically applicable measure suitable for the analysis of textual AI responses. Adding a constant in the denominator allows for the stability of the metric and the interpretation of the results in situations of low redundancy.

Such a formulation allows for a clear and theoretically grounded interpretation of the IEM metric as a measure of the information ratio of signal-to-noise in AI-generated responses, while acknowledging the goal-directed nature of human-AI communication.

The proposed information efficiency (IEM) metric can be further interpreted as a semantic and pragmatic reinterpretation of the fundamental principles of classical information theory. According to Shannon's original formulation [10], communication efficiency follows from the ratio between the useful signal and the noise present within the communication channel. Although Shannon's model was developed for technical signal transmission systems, its basic logic based on the ratio of useful information to the total communication load remains conceptually applicable in digital, semantically rich communication environments.

In this sense, the IEM metric represents a heuristic but theoretically consistent counterpart to Shannon's channel efficiency. Figure 2 illustrates the model, adapted for textual human-AI communication. Instead of physical signal and technical noise, the focus is on the informational signal in the form of relevant and dense content and the noise in the form of redundancy, repetition and low informational response segments. This translates the classical information-theoretical relationship between signal and noise into the semantic domain of human-AI dialogue.

The chosen functional form of the information efficiency metric is additionally methodologically justified by its simplicity, scalability and interpretative clarity. The multiplicative connection of information density and relevance ensures that the information value of a response is recognized only when both conditions are simultaneously satisfied, thus avoiding a situation in which long but poorly focused responses achieve high scores. Penalizing redundancy through the denominator (1+Q) allows for normalization of the metric and its applicability to responses of different lengths and structures.

Such a formulation is consistent with the info-metric approach to reasoning under uncertainty, as developed by Amos Golan et al. [19], where information value is assessed through relationships, not absolute quantities. The information efficiency metric (IEM) thus does not seek to replace classical evaluation metrics, but rather to offer a complementary, theoretically grounded measure of

information efficiency, particularly suitable for the analysis of AI-mediated communication within the contemporary infosphere, as conceptualized by Luciano Floridi [5].

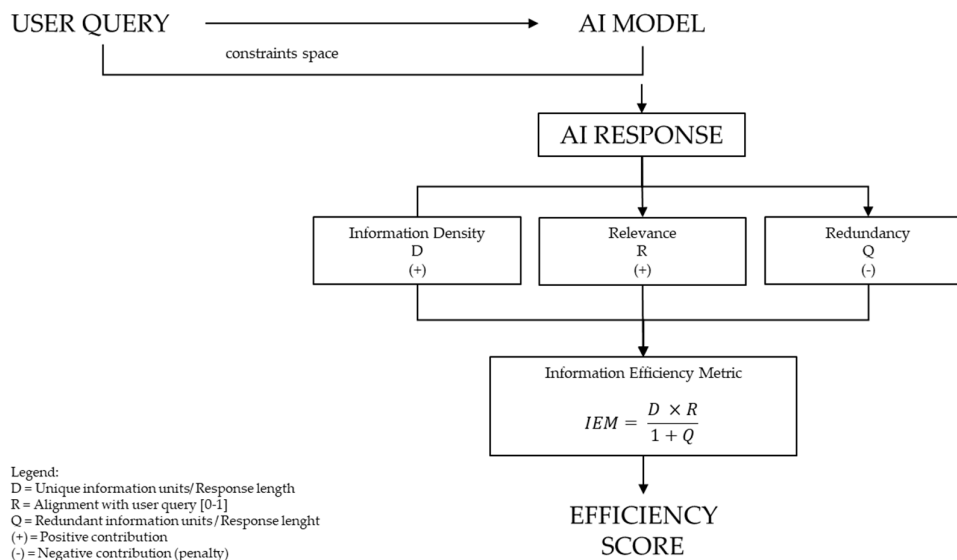


Figure 2. Information efficiency model.

4. Materials and Methods

Human–AI communication cannot be understood well enough if we observe it only through the technical characteristics of the systems in which it takes place and the subjective assessments of the responses given by the AI. The research aims to broaden the analytical focus to encompass the content structure of AI-generated responses. Such an approach requires a solidly theoretically grounded methodological framework that must at the same time be empirically operational. Methodologically, the work combines theoretical grounding in information theory and info-metrics with an experimental quantitative approach.

The research is designed as quantitative and experimental, with the aim of empirically validating the info-metric model developed for measuring information efficiency. The experimental character stems from the controlled conditions in which AI responses are generated, while the quantitative approach stems from the way in which information efficiency is operationalized, through dimensions that are clearly defined and measurable (D, R, Q). The unit of analysis is an individual AI response to a user query, treated as a unique information phenomenon that cannot be reduced to a simplistic “good” or “bad” judgement, but possesses its own distinct structure. The research sample consists of responses to predefined user queries generated by two large language models that were selected as representative examples of recent generative AI systems. To encompass the most diverse communication functions of AI systems, three types of communication queries were used: factual, analytical, and creative, across two groups of queries: queries thematically related to this work and assumed simple everyday queries of AI system users. Figure 3 shows the overall research design from question preparation to statistical analysis and interpretation.

The research was conducted from December 2025 to January 2026. Data collection was carried out on January 3, 2026, and coding and analysis in the week after.

The research used two state-of-the-art LLMs: ChatGPT version 5.2 (OpenAI) and Claude Opus version 4.5 (Anthropic). Both models were selected based on their wide application in professional and academic contexts, public availability, comparable capabilities, and representativeness of different approaches to text generation. All responses were generated with default model settings, without specific prompts or additional instructions, to ensure standardized comparison.

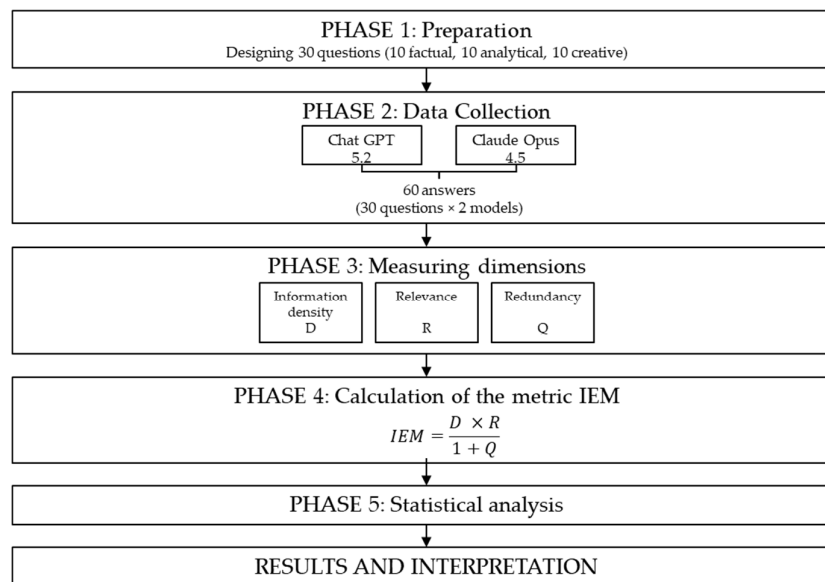


Figure 3. Research design diagram.

Data collection was conducted through a clearly defined procedure. In the first step, a set of standardized user questions was created, designed to represent the desired types and groups of communication queries, to be clear and unambiguous, and comparable between AI models. The authors wanted the questions to capture different patterns of use of generative AI systems. The question formulation was intended to reduce the possibility of ambiguous understanding of the query and to enable structured interpretation by different AI models. This approach is in line with recent findings showing that the way interactions are structured and how queries are formulated for the AI system (“prompt guidance”) has a major impact on the quality of responses of conversational AI systems in question-asking tasks [31]. Their research showed that message recommendations that structure and guide the conversation with the AI model/system can positively affect the efficiency of human–AI communication, which further justifies our approach of controlled and standardized user queries in this research.

The empirical sample consists of responses created by two large language models, which were selected as representative examples, widely used in information, education and business environments: ChatGPT 5.2 (OpenAI) and Claude Opus 4.5 (Anthropic). For each model, an equal number of responses were generated for each group and question type, ensuring sample balance and enabling comparison of results across models and task types.

A total of 30 questions were designed, systematically arranged into three categories (n=10 per category): factual questions that require retrieval of specific verifiable information, analytical questions that require thinking, synthesis and interpretation, and creative questions that require generative thinking and open-ended responses.

Factual questions:

- What does it mean that an answer is “clear”?
- What is “redundancy” in a text?
- What does “relevant” mean in an answer?
- What is a “summary”?
- In simple terms, what is “information density”?
- How many minutes are in one hour?
- What is 12×8 ?
- What does “PDF” mean?
- What is a password?

- What is a reminder?

Analytical questions:

- Why can a very long answer be a problem?
- How can you tell an answer is “wandering off topic”?
- Why does repetition reduce usefulness?
- How would you shorten an answer but keep it accurate?
- Why is relevance sometimes more important than “explaining everything”?
- How do you write a good AI prompt in one sentence?
- How do you choose daily priorities?
- Why is it good to double-check an AI answer?
- How can you tell an AI answer is too general?
- How do you shorten a text that is too long?

Creative questions:

- Create a metaphor for an answer “full of noise.”
- Write an 8 – 12 word slogan about “less, but better.”
- Write a two-sentence mini-story about an overly long answer.
- Write a short rhyme about relevance.
- Give a one-sentence rule for a good answer.
- Write a short text to a friend saying you’ll be 10 minutes late.
- Create a title for a note about a healthier lunch.
- Write one motivational sentence for studying.
- Suggest a cheap weekend idea.
- Write a one-sentence “about you” line for a profile.

Data collection was conducted on January 3, 2026. Each of the 30 questions was asked to each of the two AI models once, in separate sessions without the context of the previous questions. A total of 60 responses (30 questions × 2 models) were collected. The questions were asked via the ChatGPT and Claude web interfaces, using standard settings without additional prompts or context. Each response was saved in a Microsoft Excel (version 2511) file along with metadata (AI model, question group, question type, generation date). The quality of the collected data was ensured by standardized question formulation, the same order of asking questions for both models, collecting responses during the same time period, and no unavailability of the AI model during the response collection period.

The measurement of the information efficiency of the AI model is based on a conceptual info-metric model developed as part of this work. The model includes three dimensions: information density, relevance, and redundancy, which together describe the information model of the AI response. Information density (D) refers to the ratio of information content to response length, reflecting the AI model’s ability to produce meaningful information. Relevance (R) describes the extent to which the produced response is consistent with the user’s question and communication goal. Redundancy (Q) indicates how much of the content (repetitions, digressions, overgeneralized parts of the response) does not contribute to the desired communication goal. Together, the three dimensions form the information efficiency metric (IEM), which provides a unique quantitative evaluation of human–AI communication. This approach is in line with recent research that highlights the need for multidimensional measures of human–AI collaboration effectiveness [32]. The authors’ approach combines automated analysis of structural aspects (information density, redundancy) with human evaluation of contextual aspects (relevance), thus enabling a comprehensive assessment.

The information efficiency measurement process combined automated procedures and human evaluation. Automated analysis is used to measure structural aspects of responses, while human judgment was applied to assess semantic and pragmatic dimensions, particularly when it comes to relevance. Coding is carried out by one of the authors according to predefined criteria, and a second researcher checks the application of coding rules, where any disagreements are resolved through

discussion and consensus. Applying such a procedure allows for a reduction in subjectivity, while maintaining a high level of interpretive sensitivity necessary for analyzing the meaning of an individual response.

Each AI response was coded according to predefined criteria. Information units (I_u) were defined as semantically unique elements that provide new information relevant to the user's query. Each unit corresponded to a single meaningful piece of information that could be true or false independently of other parts of the response. Relevant units (I_{rel}) were identified as those that directly answered the question posed. Redundant units (I_{red}) included repetitions, overgeneralizations, and content that did not contribute to the answer to the query.

Coding rules:

1. Segmentation: Responses were segmented into sentences and items (clauses). Each item that conveyed a different claim was considered a candidate for a new information unit.
2. Uniqueness criterion: A unit was coded as unique (I_u) if it introduced information that was not previously provided in the response, even if expressed in different words. Different styles of expressing the same content were coded as one unit. Example: "AI is a computer program" and "Artificial intelligence is a software system" are considered the SAME information unit because they convey identical information/logic.
3. Relevance criterion: An information unit is coded as relevant (I_{rel}) if it directly answers the user's query, provides the necessary context for understanding the response, and contributes to the fulfillment of the communication goal. Units that wander into related topics or provide information that does not contribute to the goal are coded as irrelevant.
4. Redundancy criterion: A unit is coded as redundant (I_{red}) if it repeats information previously provided in the response, rephrases the query without adding new information, contains overgeneralized statements that do not add specific content (e.g., "This is an important topic"), or includes unnecessary details that do not improve understanding. What is important to note is that an information unit can be both unique (I_u) and relevant (I_{rel}), but it cannot be both relevant and redundant.

Coding was conducted in two steps. The first researcher coded all 60 AI responses, and the second researcher checked the coding results and the application of the rules. Since the check was conducted as a verification of the application of the rules instead of independent double coding, quantitative measures of inter-coder reliability were not calculated, which is a methodological limitation.

The calculation of dimensions was carried out according to Formulas (1), (2) and (3) defined in Section 3.2. The length of the answer (L) was measured by the number of words. All calculations were performed in Microsoft Excel.

In accordance with the stated research objective, research questions were set that were precisely formulated according to the PICO model, based on which the following hypotheses were formulated:

- **H1:** Information density (D) and relevance (R) have a positive effect on information efficiency (IEM), while redundancy (Q) has a negative effect on information efficiency (IEM).
- **H2:** The variability of information efficiency (IEM) between AI models is greater for factual questions than for analytical and creative questions.
- **H3:** Responses to creative questions will have a statistically significantly higher level of redundancy (Q) compared to responses to factual and analytical questions.
- **H4:** Responses to factual questions achieve higher information efficiency (IEM) compared to responses to analytical and creative questions.

Descriptive statistics (arithmetic mean, median, standard deviation, minimum and maximum) were calculated for all variables, while the normality of the distribution was tested with the Shapiro-Wilk test. Due to deviations from normal distribution, differences between AI models were analyzed with the Mann-Whitney U test, and differences between question types with the Kruskal-Wallis H test. Multivariate regression analysis was performed with the information efficiency metric (IEM) as

the dependent variable and information density (D), relevance (R) and redundancy (Q) as predictors, using the least squares method. All analyses were performed in Microsoft Excel and IBM SPSS Statistics.

The research questions and hypotheses are conceptually and empirically related to each other. IP1 and IP4 examine differences in information efficiency with respect to the type and complexity of user questions and are operationalized by hypotheses H2–H4, while IP2 focuses on differences between large language models (H2). IP3 analyzes the contribution of information density, relevance, and redundancy to explaining variability in information efficiency, which is operationalized by hypothesis H1.

The research does not involve human participants or the processing of personal data and is based solely on the analysis of text responses from publicly available AI systems, generated through their official web interfaces using insensitive, purposefully formulated queries.

Methodological limitations include the limited number of questions and responses, the small number of models analyzed, and the manual coding of information units, which introduces an element of subjectivity and limits scalability. Nevertheless, the proposed info-metric framework allows for an empirically based comparative evaluation of the information efficiency of AI responses.

5. Results

In this study, a total of 60 AI responses generated on January 3, 2026, by two AI models (ChatGPT 5.2 and Claude Opus 4.5) were analyzed. The sample was evenly distributed by AI model (30 responses per model), question type (factual, analytical, and creative), and content type (exploratory and everyday questions). For each response, the underlying textual variables were coded, and the info-metric indicators were derived: information density (D), relevance (R), redundancy (Q), and information efficiency (IEM).

Table 3 shows the descriptive statistics of the analyzed variables. Response length (L) showed high variability ($M = 81.4$; $SD = 91.37$), with a pronounced asymmetry of the distribution. Similar asymmetry was observed for the variables information density (D), redundancy (Q) and information efficiency (IEM).

Table 3. Overall descriptive statistics of variables ($N = 60$) (rounded to 3 decimal places).

Variable	N	Min	Max	Mean	Med	St dev
Response length (L)	60	1	276	81.4	24.5	91.366
Information units (Iu)	60	1	16	5.350	3.5	4.364
Relevant information units (I rel)	60	1	12	4.800	3.5	3.511
Redundant information units (I red)	60	0	7	0.783	0	1.519
Information density (D)	60	0.030	1	0.137	0.1	0.142
Relevance (R)	60	0.333	1	0.940	1	0.142
Redundancy (Q)	60	0	0.13	0.008	0	0.020
Information efficiency (IEM)	60	0.029	1	0.131	0.1	0.143

The average relevance of the responses is $R = 0.94$, with a median of 1, while the average redundancy is $Q = 0.008$, with a median of 0. The values of the information efficiency metric (IEM) are distributed in a wide range ($Min = 0.029$; $Max = 1$), with a moderate average value ($M = 0.131$; $SD = 0.143$). The distribution of the obtained values indicates significant heterogeneity in the structure of the AI responses, with differences in length, ratios of information components and overall information efficiency.

Table 4 shows descriptive statistics by AI model ($N = 30$ per model). The Claude model generated longer responses on average than ChatGPT and achieved higher average values of information density (D) and information efficiency (IEM), with greater variability. Relevance (R) is high for both models, while redundancy (Q) is low and comparable.

Table 4. Descriptive statistics by AI model (M ± Sdev) (N = 30 per model).

AI model	N	Response length (L)	Information density (D)	Relevance (R)	Redundancy (Q)	Information efficiency (IEM)
Chat GPT	30	64.033 ± 60.800	0.105 ± 0.051	0.944 ± 0.158	0.007 ± 0.023	0.097 ± 0.051
Claude	30	98.767 ± 112.540	0.171 ± 0.190	0.936 ± 0.125	0.008 ± 0.016	0.165 ± 0.192

Table 5 shows descriptive statistics by question type (N = 20 per type). Factual responses show the greatest variability in response length and derived measures and achieve the highest average values of information density (D) and information efficiency (IEM). Analytical responses are on average the longest and achieve the highest relevance (R) with the lowest redundancy (Q). Creative responses are the shortest, with moderate values of information efficiency and slightly higher average redundancy.

Table 5. Descriptive statistics by question type and model (M ± SDev) (N = 20 per question type).

Question type	N	Response length (L) (Mean ± SDev)	Information density (D) (Mean ± SDev)	Relevance (R) (Mean ± SDev)	Redundancy (Q) (Mean ± SDev)	Information efficiency (IEM) (Mean ± SDev)
factual	20	102.95 ± 99.566	0.171 ± 0.224	0.875 ± 0.166	0.010 ± 0.014	0.162 ± 0.229
analytical	20	118.200 ± 98.613	0.101 ± 0.067	0.995 ± 0.020	0.004 ± 0.006	0.101 ± 0.068
creative	20	23.050 ± 28.037	0.141 ± 0.072	0.950 ± 0.163	0.010 ± 0.031	0.131 ± 0.072

Table 6 shows the descriptive statistics of the info-metric indicators by combination of AI model and question type (N = 10 per combination). For ChatGPT, information density (D) and information efficiency (IEM) show relatively uniform values across question types, with the highest relevance (R) recorded for analytical questions. Redundancy (Q) remains low across all categories.

Table 6. Descriptive statistics by question type model (M ± Sdev) (N = 10 per question type).

AI model	Question type	N	Information density (D) (Mean ± SDev)	Relevance (R) (Mean ± SDev)	Redundancy (Q) (Mean ± SDev)	Information efficiency (IEM) (Mean ± SDev)
ChatGPT	factual	10	0.101 ± 0.047	0.900 ± 0.174	0.038 ± 0.008	0.092 ± 0.053
ChatGPT	analytical	10	0.094 ± 0.054	1.000 ± 0.000	0.006 ± 0.007	0.094 ± 0.054
ChatGPT	creative	10	0.119 ± 0.053	0.933 ± 0.211	0.013 ± 0.040	0.106 ± 0.051
Claude	factual	10	0.242 ± 0.304	0.851 ± 0.161	0.016 ± 0.017	0.213 ± 0.311
Claude	analytical	10	0.108 ± 0.081	0.991 ± 0.029	0.003 ± 0.003	0.108 ± 0.081
Claude	creative	10	0.163 ± 0.084	0.967 ± 0.105	0.007 ± 0.021	0.155 ± 0.083

To test hypothesis H1, a multiple linear regression analysis was conducted, with the information efficiency metric (IEM) defined as the dependent variable, while information density (D), relevance (R) and redundancy (Q) were included as independent variables. The results of the regression analysis showed that the model was statistically significant ($F(3,56) = 10910.93$; $p < 0.001$). The coefficient of determination is $R^2 = 0.998$, while the adjusted coefficient of determination is Adjusted

$R^2 = 0.998$, indicating that the model explains 99.8% of the variance of the information efficiency metric.

Analysis of the individual regression coefficients (Table 7) showed that information density has a strong and statistically significant positive effect on the information efficiency metric ($B = 0.990$; $\beta = 0.979$; $t = 176.22$; $p < 0.001$). Relevance also has a positive and statistically significant effect on the information efficiency metric ($B = 0.076$; $\beta = 0.075$; $t = 9.34$; $p < 0.001$). In contrast, redundancy has a negative and statistically significant effect on the information efficiency metric ($B = -0.542$; $\beta = -0.075$; $t = -9.38$; $p < 0.001$). **H1 is fully confirmed.**

Table 7. Analysis of individual regression coefficients (H1).

Model	Coefficients ^a				Collinearity Statistics		
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Tolerance	VIF
	B	Std. Error	Beta				
(Constant)	-0.073	0.008		-9.101	0.000		
1 Information density (D)	0.990	0.006	0.979	176.217	0.000	0.989	1.011
Relevance (R)	0.076	0.008	0.075	9.344	0.000	0.473	2.113
Redundancy (Q)	-0.542	0.058	-0.075	-9.379	0.000	0.477	2.097

Analyzing hypothesis H2, comparing descriptive indicators, factual questions show the highest variability of the effectiveness metric (IEM) [33]. The variance of factual questions (0.052) is significantly higher than that of analytical and creative questions (0.005), while the standard deviation of factual questions (0.229) is many times higher than that of the other two question types. Furthermore, the range of values of the information effectiveness metric for factual questions (0.965) significantly exceeds the range for analytical (0.205) and creative questions (0.295).

Levene's test for homogeneity of variance (Table 8) showed a statistically significant difference in the variability of the information effectiveness metric between question types ($F(2,57) = 3.85$, $p = 0.027$). These results provide descriptive and inferential **support for hypothesis H2**, according to which the variability of the AI model information efficiency metrics higher for factual than for analytical and creative questions.

Table 8. Levene's Test of Homogeneity of Variances (H2).

Tests of Homogeneity of Variances					
		Levene Statistic	df1	df2	Sig.
Information efficiency (IEM) IEM=(D×R)/(1+Q)	Based on Mean	3.846	2	57	0.27
	Based on Median	1.254	2	57	0.293
	Based on Median and with adjusted df	1.254	2	22.799	0.304
	Based on trimmed mean	2.089	2	57	0.133

Descriptive analysis of research hypothesis H3 showed a pronounced positive asymmetry and leptokurtic distribution of the redundancy variable (Q), especially for creative questions. Normality tests (Shapiro–Wilk) confirmed a significant deviation from normal distribution in all groups ($p < 0.001$). Visual inspection of the histograms and Q–Q plots further confirmed the presence of rare but extreme redundancy values, especially for creative responses. Due to the violated assumption of normality, the non-parametric Kruskal–Wallis test was used to test for differences between question types.

The Kruskal–Wallis test (Table 9) did not show statistically significant differences in redundancy (Q) between question types, $H(2) = 5.22$, $p = 0.074$. Thus, **H3** about a higher level of redundancy in creative answers, **was not confirmed**.

Table 9. Kruskal-Wallis test (H3).

	Test Statistics a,b		
	Kruskal-Wallis H	df	Asymp. Sig.
Redundancy (Q)	5.217	2	0.074

a. Kruskal Wallis Test
b. Grouping Variable: Q_type_num

A descriptive analysis of the research hypothesis H4 showed that answers to factual questions achieve higher information efficiency ($M = 0.162$) compared to analytical ($M = 0.101$) and creative questions ($M = 0.131$). Based on the comparison of mean values, the hypothesis **H4** that factual questions achieve higher information efficiency compared to other question types **was confirmed**.

Figure 4 shows a conceptual research model explaining the information efficiency of responses generated by artificial intelligence in human–AI interactions. The model integrates three info-metric indicators: information density (D), relevance (R), and redundancy (Q), as direct predictors of the information efficiency metric (IEM). In line with hypothesis H1, information density and relevance have a positive, while redundancy has a negative effect on efficiency.

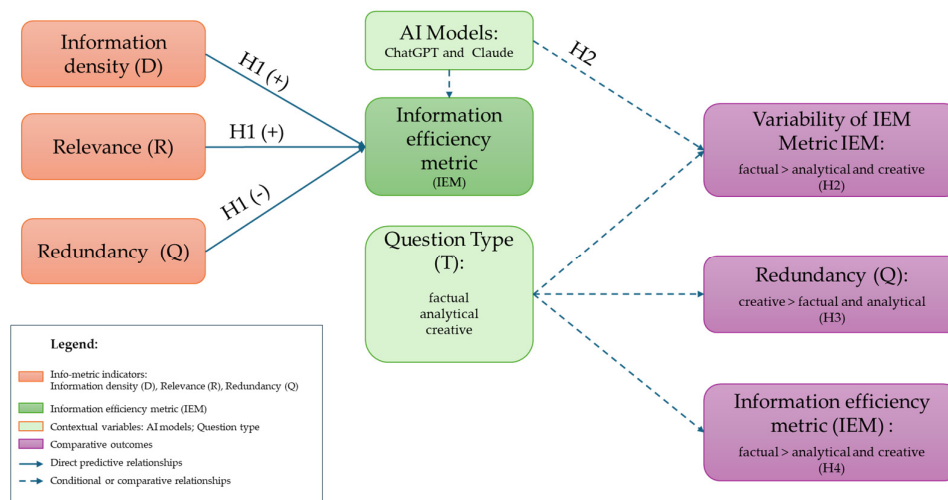


Figure 4. A conceptual model of the information efficiency of AI responses.

The model additionally includes contextual variables, namely AI models (ChatGPT and Claude) and user question type (factual, analytical, creative), which condition the variability and comparative outcomes of information efficiency. Dashed arrows indicate relationships related to hypotheses H2–H4, including differences in the variability of information efficiency metrics between AI models, differences in the level of redundancy by question type, and differences in the average information efficiency of responses. The solid arrows in the model indicate direct predictor relationships, while the dashed arrows represent conditional and comparative relationships. The model provides an integrated analytical framework for empirically assessing the information efficiency of AI responses in different interaction contexts.

6. Discussion

6.1. Interpretation of Results

The results confirm that information density (D) and relevance (R) are key positive predictors of the information efficiency metric (IEM), while redundancy (Q) has a negative effect, confirming the assumption that efficient information transfer depends on the optimal ratio of useful content to semantic redundancy. This pattern is consistent with classical information theory, according to which an increase in noise reduces communication efficiency, while strengthening the useful signal reduces uncertainty [10].

The high average relevance ($R = 0.94$) can be interpreted through a semantic extension of the Shannon–Weaver model. While Shannon (1948)[10] focused on the technical aspect of transmission, Weaver [34] highlighted the semantic problem of message interpretation. In human–AI communication, relevance operationalizes this problem by measuring the congruence of the AI response with the user’s intention. In this context, noise appears as semantic excess in the form of repetition or irrelevant content, which supports the need for a semantic extension of the information framework [17].

The following research questions were precisely formulated according to the PICO model. Each is structured so that it can be used for a systematic literature review or empirical research:

- IP1: How does the information efficiency of AI responses differ with respect to the type of user question (factual, analytical, creative)?
- IP2: Are there significant differences in information efficiency between different major language models?
- IP3: To what extent do information density, relevance, and redundancy contribute to explaining the variability of the efficiency index?
- IP4: How does the information efficiency (IEM) of AI responses change with respect to the level of complexity of the user question?

In line with IP3, the results of the regression analysis **fully confirm hypothesis H1**, with the extremely high explained variance ($R^2 = 0.998$) indicating a strong association between information structure and AI system efficiency and confirming IEM as a robust integrative metric. The analysis of variability by question type (IP1, IP2) provides **support for hypothesis H2**, showing greater variability in IEM for factual questions, suggesting less stability of performance in factually demanding contexts.

Hypothesis **H3 is not confirmed**. Although creative responses show asymmetric redundancy distributions with rare extreme values, there are no statistically significant differences between question types. Conversely, the results related to **IP4 confirm hypothesis H4**, according to which factual questions achieve a higher average value of IEM compared to analytical and creative questions.

In conclusion, the findings confirm that IEM represents a stable systematic measure of the balance between information density, relevance and redundancy, while the user question type acts as a contextual moderator of the stability of performance. These results have direct implications for the design and evaluation of AI systems aimed at efficient information processing.

6.2. Discussion of Results from the Perspective of Previous Studies

The greater variability of the information efficiency metric (IEM) in factual questions indicates that such tasks more strongly differentiate the performance of AI models, since clear constraints on accuracy and relevance reinforce differences in the ability to optimally distribute information, in accordance with the info-metric principle of entropy maximization under given constraints [19]. At the same time, the results point to the limitations of conventional benchmark-approaches in capturing user-perceived differences in open-ended dialog tasks, thus emphasizing the need for interaction-sensitive evaluation frameworks [35]. Contrary to expectations, creative responses do not show a higher level of redundancy, which suggests that semantic diversity does not necessarily represent

informational redundancy, but a plurality of relevant solutions [36]. The higher average IEM value for factual responses further highlights the role of information economy, whereby IEM can be viewed as a relational and contextual property of human–AI interaction, rather than a purely technical characteristic of the model [1]. In line with the HCI perspective and recent findings on the perception of AI communication quality [37,38], these results confirm the information efficiency metric (IEM) as a complementary evaluation dimension that connects information theory, info-metric approaches, and contemporary research on human–AI interaction [18].

6.3. Methodological Limitations and Future Research

The study has several methodological limitations that should be considered when interpreting the findings. The analysis is based on a relatively small sample of 60 responses, which was sufficient to test hypotheses and detect medium-to-large effects, but larger samples would allow for more robust interaction analysis and better generalizations. Only two large language models (ChatGPT 5.2 and Claude Opus 4.5) were tested, while generalizing the findings to other models requires testing on a wider range of LLMs, comparing different versions of the same models, and examining models of different sizes and architectures. Manual coding of information units introduces an element of subjectivity, is time-consuming, and makes scaling to large data sets difficult. Furthermore, the 30 questions, although carefully designed, represent a limited sample of possible user query types, domains, and levels of complexity within each category.

Based on the identified limitations, we suggest three main directions for future research. First, extending to a wider range of LLMs, open-source and proprietary models, and longitudinally tracking changes in performance across versions. Second, developing NLP-based methods for automatic identification of information units, comparison of automatic measures with human estimation, and creation of reference datasets for model training and evaluation. Third, domain-specific evaluation of info-metric patterns in specific domains, with the development of domain-adapted dimensions of relevance and examination of the influence of domain expertise on performance evaluation.

6.4. Practical Implications

The proposed info-metric framework has practical implications for three groups: (1) users of AI systems, by enabling more critical evaluation of AI responses based on the ratio of useful content to semantic redundancy; (2) development teams, by offering a complementary evaluation tool for optimizing models according to information economy; and (3) the research community, by providing the first systematic attempt to operationalize info-metric principles in human–AI communication evaluations.

The obtained results have important empirical implications for the evaluation of AI-mediated communication. The positive effect of information density and relevance, along with the negative effect of redundancy, confirms that information efficiency can be reliably measured by quantitative structural metrics, thus going beyond evaluations based solely on subjective assessments or accuracy of responses [23,25]. The greater variability of efficiency in factual tasks further confirms the need for task-sensitive evaluation protocols. Furthermore, the finding that creative responses do not show increased redundancy suggests that the variety of expressions in open-ended tasks represents relevant semantic dispersion, and not necessarily information redundancy. Overall, the results support the understanding of information efficiency as a relational and contextual property of human–AI interaction [39].

7. Conclusions

In this research, an empirically validated info-metric model for quantifying information efficiency in human–AI communication is developed. The model integrates three theoretically grounded dimensions: information density (D), relevance (R), and redundancy (Q) synthesized into

a single information efficiency metric (IEM). Empirical validation conducted on 60 AI responses confirms that the model successfully distinguishes efficiency by task type, with factual questions achieving the highest efficiency and the highest discriminativeness among models. The framework offers a task-sensitive, theoretically grounded approach to evaluating human–AI communication beyond traditional accuracy metrics or subjective quality assessments.

Author Contributions: Conceptualization, M.R.; methodology, M.R.; software, M.B.; validation, M.R., M.B.; formal analysis, M.R., M.B.; investigation, M.R.; resources, M.R.; data curation, M.R.; writing—original draft preparation, M.R.; writing—review and editing, L.L.; visualization, M.R., M.B.; supervision, L.L.; project administration, L.L.; funding acquisition, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: The publication of this article was made possible by a grant from the University North (UNIN-DRUŠ-25-1-4). <https://www.croris.hr/projekti/projekt/13201> accessed on 3 February 2026.

Data Availability Statement: Data will be made available on request by the corresponding author.

Acknowledgments: The publication of this paper was made possible through the University North funds, intended to support the scientific research project “Research perspectives of the Design Thinking method in the fields of social sciences and interdisciplinary sciences”, to which these authors express their gratitude.

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

Abbreviations

The following abbreviations are used in this manuscript:

IEM	Information Efficiency Metric
LLM	Large Language Model

References

1. Guzman, A.L.; Lewis, S.C. Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media Soc.* 2020, 22, 70–86. <https://doi.org/10.1177/1461444819858691>
2. Guzman, A. Voices in and of the Machine: Source Orientation toward Mobile Virtual Assistants. *Computers in Human Behavior* 2019. <https://doi.org/10.1016/j.chb.2018.08.009>
3. Mirek-Rogowska, A., Kucza, W., Gajdka, K., 2024. AI in Communication: Theoretical Perspectives, Ethical Implications, and Emerging Competencies. *CT* 16–29. <https://doi.org/10.34135/communicationtoday.2024.Vol.15.No.2.2>
4. Shneiderman, B. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction* 2020, 36, 495–504. <https://doi.org/10.1080/10447318.2020.1741118>
5. Floridi, L. *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford University Press: Oxford, UK, 2014.
6. Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; Schafer, B.; Valcke, P.; Vayena, E. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds Mach.* 2018, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
7. Luger, E.; Sellen, A. “Like Having a Really Bad PA”: The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 5286–5297. <https://doi.org/10.1145/2858036.2858288>
8. Aftab, O.; Cheung, P.; Kim, A.; Thakkar, S.; Yeddanapudi, N. *Information Theory and the Digital Revolution*. Unpublished manuscript, Massachusetts Institute of Technology, Cambridge, MA, USA, 2001.
9. Banks, J.; Bowman, N.D. Avatars Are (Sometimes) People Too: Linguistic Indicators of Parasocial and Social Ties in Player–Avatar Relationships. *New Media & Society* 2016, 18, 1257–1276

10. Shannon, C.E. A Mathematical Theory of Communication. *Bell System Technical Journal* 1948, 27, 379–423, 623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
11. Gao, J.; Galley, M.; Li, L. Neural Approaches to Conversational AI. *Foundations and Trends® in Information Retrieval* 2019, 13, 127–298
12. Mahmud, B.; Hong, G.; Fong, B. A Study of Human–AI Symbiosis for Creative Work: Recent Developments and Future Directions in Deep Learning. *ACM Transactions on Multimedia Computing, Communications, and Applications* 2023, 20, 47:1–47:21. <https://doi.org/10.1145/3542698>
13. Hassani, H.; Silva, E.S.; Unger, S.; TajMazinani, M.; Mac Feely, S. Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future? *AI* 2020, 1, 143–155. <https://doi.org/10.3390/ai1020008>
14. Cabitza, F.; Campagner, A.; Fregosi, C.; Cameli, M.; Gallazzi, E.; Sconfienza, L.M.; Tontini, G.E. Five Degrees of Separation: Investigating the Unexpected Potential of Displaced Human–AI Collaboration Protocols for Apter AI Support. *Proceedings of the ACM on Human-Computer Interaction* 2025, 9. <https://doi.org/10.1145/3757601>
15. Epstein, Z.; Hertzmann, A.; Investigators of Human Creativity. *Art and the Science of Generative AI: Understanding Shifts in Creative Work Will Help Guide AI’s Impact on the Media Ecosystem*. *Science* 2023, 380, 1110–1111. <https://doi.org/10.1126/science.adh445>
16. Bajić, D. Information Theory, Living Systems, and Communication Engineering. *Entropy* 2024, 26, 430. <https://doi.org/10.3390/e26050430>
17. Floridi, L. *The Philosophy of Information*; Oxford University Press: Oxford, UK, 2011.
18. Golan, A. *Foundations of Info-Metrics: Modeling, Inference, and Imperfect Information*. Oxford University Press: Oxford, UK, 2018.
19. Golan, A. On the State of the Art of Info-Metrics. In *Uncertainty Analysis in Econometrics with Applications*; Springer: Berlin, Heidelberg, Germany, 2013; pp. 3–15.
20. Anbalagan, B.; Valliyammai, C. Information Entropy Based Event Detection during Disaster in Cyber-Social Networks. *Journal of Intelligent & Fuzzy Systems* 2019, 36, 1–12. <https://doi.org/10.3233/JIFS-169959>
21. Kumar, A.; Rana, S.; Shilton, A.; Venkatesh, S. Human–AI Collaborative Bayesian Optimisation. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*; Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; Oh, A., Eds.; 2022.
22. Amodei, D.; Olah, C.; Steinhardt, J.; Christiano, P.; Schulman, J.; Mané, D. Concrete Problems in AI Safety. *arXiv* 2016, arXiv:1606.06565. <https://arxiv.org/abs/1606.06565>
23. Gehrmann, S.; Adewumi, T.; Aggarwal, K.; Aremu, A.; Bosselut, A.; et al. The GEM Benchmark: Natural Language Generation, Its Evaluation and Metrics. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2021, pp. 96–120. <https://aclanthology.org/2021.gem-1.10/>
24. Huang, L.; Yu, W.; Ma, W.; Zhong, W.; Feng, Z.; Wang, H.; Chen, Q.; Peng, W.; Feng, X.; Qin, B.; Liu, T. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Computing Surveys* 2023. <https://doi.org/10.1145/3703155>
25. Karpinska, M.; Akoury, N.; Iyyer, M. The Perils of Using Mechanical Turk to Evaluate Open-Ended Text Generation. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*; Association for Computational Linguistics: Online & Punta Cana, Dominican Republic, 2021; pp. 1265–1285. <https://doi.org/10.18653/v1/2021.emnlp-main.97>
26. Hao, X.; Demir, E.; Eysers, D. Exploring Collaborative Decision-Making: A Quasi-Experimental Study of Human and Generative AI Interaction. *Technology in Society* 2024, 78, 102662. <https://doi.org/10.1016/j.techsoc.2024.102662>
27. Rosbach, E.; Ammeling, J.; Krügel, S.; Kiessig, A.; Fritz, A.; Ganz, J.; Puget, C.; Donovan, T.; Klang, A.; Köller, M.; Bolfa, P.; Tecilla, M.; Denk, D.; Kiupel, M.; Paraschou, G.; Kok, M.; Haake, A.; de Krijger, R.; Sonnen, A.; Kasantikul, T.; Dorrestein, G.; Smedley, R.; Stathonikos, N.; Uhl, M.; Bertram, C.; Riener, A.; Aubreville, M. “When Two Wrongs Don’t Make a Right”: Examining Confirmation Bias and the Role of Time Pressure during Human–AI Collaboration in Computational Pathology. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI 2025)*; Association for Computing Machinery: New York, NY, USA, 2025. <https://doi.org/10.1145/3706598.3713319>

28. Golan, A.; Maasoumi, E. Information theoretic and entropy methods: An overview. *Econometr. Rev.* 2008, *27*, 317–328. <https://doi.org/10.1080/07474930801959685>
29. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* 2021, *372*, n71. <https://doi.org/10.1136/bmj.n71>
30. Sengar, S.S.; Hasan, A.B.; Kumar, S.; Carroll, F. Generative Artificial Intelligence: A Systematic Review and Applications. *Multimedia Tools and Applications* 2025, *84*, 23661–23700. <https://doi.org/10.1007/s11042-024-20016-1>
31. Song, J.; Ashktorab, Z.; Pan, Q.; Dugan, C.; Geyer, W.; Malone, T.W. Interaction Configurations and Prompt Guidance in Conversational AI for Question Answering in Human–AI Teams. *Proceedings of the ACM on Human-Computer Interaction* 2025, *9*. <https://doi.org/10.1145/3757486>
32. Hosawi, A.; Stone, R. Enhancing Trust in Human–AI Collaboration: A Conceptual Review of Operator–AI Teamwork. *International Journal of Advanced Computer Science and Applications* 2025, *16*, 1–15. <https://doi.org/10.14569/IJACSA.2025.0160701>
33. Chiang, W.-L.; Li, L.; Zheng, T.; Li, J.; Zhang, M.; Zhang, X.; Sheng, J.; Jin, Y.; Zhu, Z.; Chen, A.; et al. Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference. *arXiv* 2023, arXiv:2306.05685. <https://arxiv.org/abs/2306.05685>
34. Weaver, W. Recent Contributions to the Mathematical Theory of Communication. *ETC: A Review of General Semantics* 1953, *10*, 261–281
35. Zheng, L.; Chiang, W.-L.; Sheng, Y.; Zhuang, S.; Wu, Z.; Zhuang, Y.; Lin, Z.; Li, Z.; Li, D.; Xing, E.P.; Zhang, H.; Gonzalez, J.E.; Stoica, I. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. *arXiv* 2023, arXiv:2306.05685v4. <https://arxiv.org/abs/2306.05685v4>
36. Boden, M.A. *AI: Its Nature and Future*. Oxford University Press: Oxford, UK, 2016.
37. Amershi, S.; Cakmak, M.; Knox, W.B.; Kulesza, T. Guidelines for Human–AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, UK, ACM, 2019, pp. 1–13. <https://doi.org/10.1145/3290605.3300233>
38. Pan, W.; Liu, D.; Meng, J.; Liu, H. Human–AI Communication in Initial Encounters: How AI Agency Affects Trust, Liking, and Chat Quality Evaluation. *New Media & Society* 2025, *27*, 5822–5847. <https://doi.org/10.1177/14614448241259149>
39. Hancock, J.T.; Naaman, M.; Levy, K. AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. *Journal of Computer-Mediated Communication* 2020, *25*, 89–100. <https://doi.org/10.1093/jcmc/zmz022>

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