Maria Ciência: Application of Artificial Intelligence for Audience-Specific Health Communication and Knowledge Dissemination

Mariana Araújo-Pereira, Klauss Villalva-Serra Jr, Gustavo Pires-Ramos, Beatriz Sousa-Peres, Joanã Nascimento Conceição-Oliveira, Sarah Dourado Maiche, Rebeca Rebouças da Cunha Silva, Bruno de Bezerril Andrade

Submitted to: JMIR Infodemiology
on: June 10, 2025

Table of Contents

Maria Ciência: Application of Artificial Intelligence for Audience-Specific Health Communication and Knowledge Dissemination

Mariana Araújo-Pereira[1] PhD; Klauss Villalva-Serra Jr[1] MD; Gustavo Pires-Ramos[2]; Beatriz Sousa-Peres[2]; Joanã Nascimento Conceição-Oliveira[1] BSc; Sarah Dourado Maiche[1]; Rebeca Rebouças da Cunha Silva[3] BSc; Bruno de Bezerril Andrade[1] MD, PhD

[1] Instituto Gonçalo Moniz Salvador BR
[2] Universidade Federal da Bahia Salvador BR
[3] Escola Bahiana de Medicina e Saúde Pública Salvador BR

Corresponding Author:
Bruno de Bezerril Andrade MD, PhD

Instituto Gonçalo Moniz
R. Waldemar Falcão 121
Salvador
BR

Abstract

Background: Scientific misinformation remains a major barrier to effective health communication. Bridging the gap between academic research and public understanding requires tools that simplify scientific language and adapt content to diverse audiences.

Objective: This study presents Maria Ciência, a specialized GPT-based assistant for science communication. The tool supports researchers in translating peer-reviewed scientific findings through simple prompts into accessible, ethically appropriate materials tailored for children, the general public, health professionals, and policymakers.

Methods: The tool was configured using prompt engineering techniques and guided by curated reference materials on inclusive and non-stigmatizing scientific language. Materials derived from 47 public health articles resulted in 188 outputs, which were assessed by 121 evaluators using four criteria: clarity, level of detail, language suitability, and content quality.

Results: Globally, mean scores were high: clarity (4.90), language suitability (4.78), content quality (4.72), and level of detail (4.56), on a 5-point scale. Materials for children and the general public consistently achieved the highest ratings across all criteria.

Conclusions: A targeted comparison with the base large language model (ChatGPT 4o) demonstrated superior performance of Maria Ciência in contextual stability. Maria Ciência demonstrates the potential of AI-assisted tools to enhance knowledge translation and counter scientific misinformation by producing scalable, audience-specific content.

(JMIR Preprints 10/06/2025:78843)
DOI: https://doi.org/10.2196/preprints.78843

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

Maria Ciência: Application of Artificial Intelligence for Audience-Specific Health Communication and Knowledge Dissemination

Mariana Araújo-Pereira[1,2,3**], Klauss Villalva-Serra[1,2,3], Gustavo Pires Ramos Cerqueira dos Santos[1,3,4], Beatriz S Peres[1,3,5], Joanã Nascimento da Conceição Oliveira[1,3], Sarah Dourado Maiche[1,3,7], Rebeca Rebouças da Cunha Silva[1,2,3,6], Bruno B Andrade[1,2,3,6,7*]

[1]Laboratório de Pesquisa Clínica e Translacional, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil.
[2]The BRIGHT lab, Multinational Organization Network Sponsoring Translational and Epidemiological Research (MONSTER) Institute, Salvador, Brazil.
[3]SYBIL Lab, Multinational Organization Network Sponsoring Translational and Epidemiological Research (MONSTER) Institute, Salvador, Brazil.
[4]Curso de Graduação em Biotecnologia, Universidade Federal da Bahia, Salvador, Brazil.
[5]Curso de Graduação em Medicina, Universidade Federal da Bahia, Salvador, Brazil.
[6]Programa de Pós-Graduação de Tecnologias em Saúde, Escola Bahiana de Medicina e Saúde Pública (EBMSP), Salvador, Brazil.
[7]Curso de Graduação em Medicina, Medicina Zarns, Salvador, Brazil.

*Correspondence: Bruno B. Andrade, MD, PhD, Laboratório de Pesquisa Clínica e Translacional, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Rua Waldemar Falcão, 121, Candeal, Salvador, Bahia 40296-710, Brazil. Phone: +55-71-3176-2264 E-mail: bruno.andrade@fiocruz.br

**Co-corresponding author: Mariana Araújo-Pereira, PhD, Laboratório de Pesquisa Clínica e Translacional, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Rua Waldemar Falcão, 121, Candeal, Salvador, Bahia 40296-710, Brazil. Phone: +55-71-3176-2264 E-mail: mariana.araujo@fiocruz.br

ABSTRACT

Background: Scientific misinformation remains a major barrier to effective health communication. Bridging the gap between academic research and public understanding requires tools that simplify scientific language and adapt content to diverse audiences. Objective: This study presents *Maria Ciência*, a specialized GPT-based assistant for science communication. The tool supports researchers in translating peer-reviewed scientific findings through simple prompts into accessible, ethically appropriate materials tailored for children, the general public, health professionals, and policymakers. Methods: The tool was configured using prompt engineering techniques and guided by curated reference materials on inclusive and non-stigmatizing scientific language. Materials derived from 47 public health articles resulted in 188 outputs, which were assessed by 121 evaluators using four criteria: clarity, level of detail, language suitability, and content quality. Results: Globally, mean scores were high: clarity (4.90), language suitability (4.78), content quality (4.72), and level of detail (4.56), on a 5-point scale. Materials for children and the general public consistently achieved the highest ratings across all criteria. A targeted comparison with the base large language model (ChatGPT 4.5) demonstrated superior performance of *Maria Ciência* in contextual stability. Conclusions: *Maria Ciência* demonstrates the potential of AI-assisted tools to enhance knowledge translation and counter scientific misinformation by producing scalable, audience-specific content.

Keywords: Science communication, custom GPT, public health, scientific literacy

INTRODUCTION

Scientific misinformation is one of the most pressing challenges of our time, with direct consequences for public trust, the implementation of health policies, and the protection of population health[1]. The COVID-19 pandemic sharply illustrated how inaccurate or distorted information can compromise collective responses, with measurable impacts on morbidity and mortality[2]. The term “infodemic”, adopted by the World Health Organization (WHO), refers to the overwhelming volume of both accurate and misleading content that undermines access to trustworthy guidance[3]. This phenomenon is amplified by digital platforms and social media, where misinformation spreads faster than corrective content. Efforts to manage infodemics have become a policy priority for global and national institutions[3–5], particularly in response to phenomena like vaccine hesitancy, which has been linked to the resurgence of diseases such as measles[6].

Addressing misinformation requires more than reactive fact-checking, it demands the proactive translation of scientific knowledge into accessible, contextually relevant communication. Effective communication must also be timely, audience-centered, and grounded in the social and cultural contexts of the target populations. However, the scientific community itself often struggles to engage effectively with non-specialist audiences. The persistence of a publication-centred academic culture, combined with time constraints and a lack of training, limits researchers' ability to participate in outreach activities [7]. While efforts to integrate science communication into academic curricula are increasing [8,9], there remains a need for structural support and practical tools to facilitate engagement beyond scholarly environments.

This article presents *Maria Ciência (https://chatgpt.com/g/g-0DzqDWMt4-mariaciencia)*: an AI-assisted platform designed to translate scientific content into tailored, accessible formats for diverse audiences. Developed using a custom GPT model [10], the tool supports researchers in generating science communication materials adapted for children, adults with low literacy, health professionals, and decision-makers. The platform was created with the goal of enhancing the reach and impact of health-related scientific information, particularly in contexts where misinformation can influence individual behaviors and public health outcomes. the approach integrates artificial intelligence with ethical oversight, thematic supervision, and practical communication strategies. In contrast to generic chatbot applications, *Maria Ciência* is grounded in bioethical principles, equity-driven design, and a commitment to cultural and educational inclusivity. By enabling the same scientific input to be transformed into multiple outputs, the platform offers a scalable and adaptable response to the challenges of misinformation and scientific inaccessibility in public health and beyond.

METHODS

Study design

This study presents the development and evaluation of *Maria Ciência*, an AI-powered assistant designed to translate peer-reviewed research into audience-specific communication products. The tool was developed with the objective of promoting inclusive, accurate, and culturally sensitive dissemination of scientific knowledge, particularly in the field of public health. The methodological approach combined AI-supported content generation, audience-specific language adaptation, and empirical evaluation based on stakeholder feedback.

Creation and structure of the tool

Maria Ciência was developed on the ChatGPT Plus platform, a commercial version of

ChatGPT that allows for the creation of customized GPTs (also known as artificial intelligence assistants) through detailed configuration of instructions, role definitions, and operational parameters. The primary function of *Maria Ciência* is to enable researchers to input scientific materials, such as peer-reviewed articles, and select the intended target audience. Based on this input, the assistant generates adapted communication materials suitable for various reader profiles, including children, the general population, health professionals, and policy decision-makers. These outputs are designed to be practical for use in educational, clinical, and public outreach contexts. ChatGPT is a generative artificial intelligence (AIGC) developed by OpenAI that uses a transformer decoder-only architecture. The GPT model used in this approach was ChatGPT 4.5, released in February 2025. This version stands out as one of the most current versions alongside the ChatGPT o4-mini version, achieving in accuracy tests a 62.5% correctness rate, the highest value in software quality tests among the versions, and the lowest hallucination rate at 37.1% [11].

Prompt configuration and operational guidelines

The *Maria Ciência* assistant was configured through prompt engineering techniques combined with a carefully curated reference documents that established linguistic, ethical, and stylistic parameters. This configuration aimed to guide the assistant in generating respectful, inclusive, and audience-appropriate outputs. Training the AI assistant with carefully selected documentation greatly improves the accuracy and relevance of its responses [12,13]. In this study, we used reference materials on scientific dissemination, inclusive languages, and health dictionaries to fine-tune the model. Of note, the tool was initially tested in Brazilian Portuguese.

The assistant was assigned the explicit role of a Specialized Science Communicator, with advanced knowledge in public health, immunology, infectious and chronic diseases, and health communication. Its core objective was to translate complex scientific knowledge into formats that could be readily understood by diverse audiences. The interaction protocol included three primary steps:

A. Understanding user needs: prompting users to specify the target audience (children, general public, health students, or health managers);
B. Continuous engagement: encouraging deeper interaction through follow-up questions;
C. Final content generation: producing the final adapted text aligned with the audience's profile.

Content adaptation guidelines were explicitly defined for each audience segment:

A. Children: playful and narrative-driven writing inspired by Writing for Young Minds, incorporating contextual summaries and storytelling techniques;
B. General public: simplification of scientific concepts with actionable health information, drawing from accessible Brazilian sources such as Superinteressante and Profissão Biotec;
C. Health students: simplified explanations while maintaining technical terminology, with emphasis on key health concepts;
D. Health professionals and managers: structured summaries including population characteristics, methodological overviews, and actionable public health recommendations (typically five suggested policy improvements).

Additionally, for social media content, the assistant adapted outputs for platforms such as Instagram and LinkedIn, ensuring appropriate tone, hashtags, and visual alignment for each context.

Training materials

Among the sources used were the *"Na ponta das línguas: pequeno glossário para apoiar o enfrentamento do estigma e da discriminação"* carried out with the support of the Brazilian Ministry of Health, which offers terminology to reduce stigma in health communication[14], as well as the "UNAIDS Terminology Guidelines"[15], which provide recommendations for respectful and accurate language related to HIV and global health. The "Global TB Dictionary" was used to ensure technical accuracy in tuberculosis-related content, and also the "UNICEF Terminology Dictionary" for the protection of children regarding matters of a sexual nature [16]. To support outputs for younger audiences, the tool's language was informed by resources such as "*Writing for Young Minds"* and adapted scientific texts like those published in Frontiers for Young Minds [17]. Additional materials, including guides to science communication [18,19] and examples of accessible writing from Brazilian science outreach initiatives such as "Profissão Biotec" [20], were used to calibrate tone, clarity, and structure across all outputs.

The training process emphasized the use of non-stigmatizing and inclusive language, minimization of unnecessary technical jargon, clear structuring of information according to literacy level, and contextual sensitivity to cultural and ethical dimensions of health communication. An additional focus was placed on ensuring equitable representation and fairness in responses across diverse population groups, particularly those historically marginalized in public health communication. The outputs generated by *Maria Ciência* included short narratives and analogies for children, simplified articles for the general population, technical summaries for health professionals, strategic briefs for health managers, and communication materials designed for social media. Additionally, *Maria Ciência* includes, in its presentation on the question bar, pre-configured prompts for users with guiding questions for using the chatbot, such as 'Translate this article for the general public', 'How to explain this concept to children', or 'Create an accessible summary for children'. Finally, the assistant was programmed to provide appropriate attribution for all generated materials, including the original scientific source, first author, journal and year of publication.

Generation of outputs for evaluation

Following the configuration of the assistant, we selected 47 peer-reviewed articles (on public health, infectious diseases and epidemiology) from our institution to serve as the training material. This approach provided the opportunity to invite the original authors of these articles to participate as evaluators in the assessment team. To minimize potential evaluation bias, the chatbot outputs for each article were generated by external collaborators who were not part of the research group and who did not have a scientific background. Using only the pre-configured guiding questions in the chatbot interface these collaborators generated four outputs per article, including "For children," "For health managers," "For social media," and "For general public," resulting in a total of 188 outputs. All outputs were generated using a dedicated user account created exclusively for this purpose, to minimize potential bias from prior sessions or unrelated model interactions.

Public evaluation of chatbot outputs

Following the generation of 188 outputs (four per article across 47 selected articles), a public evaluation process was conducted to assess the quality and appropriateness of the content produced by *Maria Ciência*. A standardized online evaluation form was developed for this purpose, structured to allow systematic feedback from diverse audiences. The evaluation process engaged five stakeholder groups: authors of the original articles, health professionals and students, social media specialists, members of the general public and health managers. The evaluation form was disseminated through social media channels to reach a broad and heterogeneous audience. In

addition, the original authors of the selected articles were invited to participate in the evaluation. This dual approach enabled the inclusion of both expert and lay perspectives in the assessment process. Participation was voluntary and anonymous.

Each participant was asked to assess selected chatbot outputs according to the following criteria:

1) **Clarity:** Is the text clear and appropriate for the intended audience?
2) **Detail:** Does the text provide sufficient and relevant information?
3) **Language Suitability:** Is the language appropriate for the literacy level and context of the intended audience?
4) **Content quality:** Does the text maintain scientific accuracy and communicative effectiveness?

Participants assigned scores on a scale from 1 (poor) to 5 (excellent) for each criterion. Additionally, the form included an open field for qualitative comments, enabling participants to provide contextual feedback on strengths, limitations, or suggestions for improvement. Qualitative feedback provided by anonymous evaluators was analyzed using thematic categorization. Comments were grouped into four predefined domains: (1) Language (clarity, accessibility, appropriateness of language use); (2) Information (accuracy, level of detail, appropriateness of content for the audience); (3) Structure (organization, narrative flow, format of the material); and (4) Proposal (whether the material complied with the intended purpose and target audience). For each domain, comments were further classified as either Criticism or Praise (for Language, Information, and Structure), or as Complies with proposal/Does not comply with proposal (for Proposal). Each comment was reviewed independently and could be assigned to multiple categories when it addressed more than one thematic domain.

Accuracy evaluation and comparison with base GPT

In addition to the public evaluation of the outputs, a focused accuracy evaluation was conducted to compare the performance of *Maria Ciência* with that of the base ChatGPT 4o model. This comparison aimed to assess whether the custom configuration of *Maria Ciência*, which incorporates training on inclusive, non-stigmatizing, and health-appropriate language, enhanced the model's ability to maintain contextual relevance and communicative precision in public health content generation. For this purpose, a selected set of questions previously answered by *Maria Ciência* was resubmitted to both *Maria Ciência* and the base ChatGPT 4o model (without the custom configuration) separately.

To minimize potential bias from session memory, priming, or model adaptation effects, the evaluations for *Maria Ciência* and GPT 4o were conducted using separate user accounts. This ensured that the comparative responses were generated independently, reducing the risk of inadvertent learning from previous interactions within the same account. In addition, to assess response stability, the same questions were submitted multiple times in different conversational sequences, allowing us to evaluate whether the models maintained contextual coherence across repeated interactions.

Responses from both models were evaluated by a team composed of healthcare professionals and undergraduate students. The evaluation criteria included four key dimensions: whether the response accurately established the context of the question; whether contextual coherence was preserved throughout the conversation; whether there was any interruption or drift from the intended context; and, if such drift occurred, whether the model was able to recover and return to the appropriate context. Each dimension was scored on a qualitative scale from 1 (poor) to 5 (excellent). The comparative analysis of results aimed to determine whether *Maria Ciência's* configuration effectively enhanced contextual accuracy and stability, thus supporting its suitability for reliable and ethically appropriate use in public health communication. The following four dimensions were used

to structure the assessment:

> **1. Establishment of a context:** Does the answer fulfill the objective of the question according to the inserted context?
> **2. Continuity of conversation without specific context of the question:** Do the following answers in the conversation with the chatbot lose the context in relation to the question?
> **3. Interruption of context:** Do the answers interrupt or stop fulfilling the context of the question?
> **4. Return to context:** Even after the interruption of the context or digression, is the ChatBot able to return to the context of the question?

The tool operates in various languages, having been tested by the developers at Maria Ciência in Portuguese, English, Spanish, Italian, and French. Moreover, since it is a GPT assistant, the platform supports more than 50 languages [21]. Regardless of the language, the generated content follows the same principles of technical configuration, linguistic curation, and thematic supervision. This version preserves the commitment to accessibility, scientific accuracy, and ethical adequacy in knowledge translation, with a view toward application in international and multilingual contexts.

Data analysis

Quantitative data from the evaluation forms were analyzed using descriptive statistics. For each evaluation criteria, means and standard deviations (SD) were calculated. Frequencies and percentages were computed for the classification of evaluator identities and for the distribution of outputs across target audiences. All analyses were conducted using the structured database generated from the stakeholder evaluations.

RESULTS

Before launching the public evaluation, we first analyzed the process of generating the 188 outputs used for assessment. The generation was conducted by external collaborators with no scientific background, using a dedicated user account created exclusively for this purpose. During this process, important differences were observed between the use of the free version and the paid version (ChatGPT Plus). In the free version, the model frequently (10%) exhibited technical limitations: it would often require questions to be reformulated and occasionally produce incomplete outputs. The average generation time per output in this version was approximately 15.2 seconds (SD: 2.3). After upgrading to the ChatGPT Plus version, performance improved substantially. The model produced responses more rapidly and with greater consistency, showing a more direct communication style. In this version, the mean generation time decreased to 8.5 seconds (SD: 1.5), and the incidence of incomplete outputs was eliminated. Despite these differences, the free version remained capable of generating the requested outputs using the predefined prompts provided by *Maria Ciência*; however, it required longer processing time and occasional manual re-submission of prompts to ensure complete responses.

Evaluator profile and distribution of reviewed outputs

The evaluation of *Maria Ciência* involved 121 responses to the form, stratified across five distinct groups, each representing a key target audience of the tool. The distribution of respondents was as follows: health professionals and students (n = 56, 44.6%), original authors of the scientific

articles (n = 33, 27.3%), members of the general population (n = 26, 21.5%), communication specialists (n = 4, 3.3%), and health managers (n = 4, 3.3%) (**Figure 1A**). Each evaluator assessed outputs generated for one or more specific target audiences, including “For children” (n = 27), “For health managers” (n = 35), “For social media” (n = 24), and “For general public” (n = 35). Among these groups, only the original article authors had prior in-depth knowledge of the scientific content being communicated. Notably, **Figure 1B** illustrates the distribution of reviewed text types according to identity profile classification, revealing clear patterns in how different stakeholder groups engaged with the audience-adapted outputs. Communication professionals and Health Managers exclusively reviewed the solely texts targeted at their respective profiles (100%), while the other evaluator profiles diversified their reviewed texts.

Among the general population evaluators, 62% assessed the “For general public” texts, while 23% provided feedback on “For children” versions, and 15% on “For health managers”. Notably, none of the general population participants reviewed the “For social media” outputs. Health professionals and students provided a broader distribution of feedback, with 44% evaluating “For health managers” outputs, 26% “For general public”, 20% “For children”, and 9% “For social media”. Finally, the article authors demonstrated a balanced engagement across all four categories: 33% reviewed “For general public” outputs, 30% “For children”, 21% “For health managers”, and 15% “For social media” (**Figure 1B**). The distribution of respondents and their evaluation profiles across different target audiences is summarized in **Table 1**.

Public evaluation of audience-adapted outputs

Each output was independently evaluated by members of the intended target audience and by other stakeholder groups, using four evaluation criteria: clarity of the text, level of detail, suitability of language for the intended audience, and overall content quality. Participants rated each criterion on a five-point scale, with 5 indicating strong agreement regarding the quality or appropriateness of the item assessed. The distribution of evaluator profiles and scoring frequencies is presented in **Table 1**.

Overall, the adapted texts were rated highly across all criteria and target audiences (**Table 1**). Texts targeting the general public received particularly strong evaluations. Among members of this audience, mean (SD) scores were 4.94 (0.25) for clarity, 4.56 (0.89) for detail, 4.62 (0.62) for language suitability, and 4.75 (0.45) for overall quality. Students and health professionals were equally enthusiastic, with scores of 5.00 (0.00) for clarity, 4.86 (0.36) for detailing, 4.93 (0.27) for language, and 4.93 (0.27) for quality. Of note, researchers who authored the original articles were somewhat more critical, assigning 5.00 (0.00) for clarity, 4.80 (0.45) for detail, 4.80 (0.45) for language, and 4.60 (0.55) for overall quality (**Table 2**).

Similarly, the child-focused outputs were well received. The general public assigned near-perfect ratings of 5.00 (0.00) for clarity, 4.83 (0.41) for detail, 4.83 (0.41) for language suitability, and 4.83 (0.41) for content quality. Students and health professionals also rated the child-focused content highly, with a means of 4.82 (0.40) for clarity, 4.64 (0.67) for detailing, 4.73 (0.65) on language suitability, and 4.73 (0.47) regarding overall quality. Again, the original authors were more reserved, assigning 4.80 (0.42) for clarity, 4.50 (0.71) for detail, 4.40 (0.84) for language, and 4.70 (0.48) for overall quality (**Table 2**).

Next, the version tailored for health managers was more harshly reviewed by managers themselves, who rated it 4.00 (1.41) for clarity, 4.00 (1.41) for detail, 4.00 (0.82) for language suitability, and 4.25 (0.96) for content quality. Students and health professionals provided comparably high scores, with 4.83 (0.48) for clarity, 4.58 (0.72) for detail, 4.62 (0.71) for language suitability, and 4.67 (0.56) for overall quality. Once again, the authors were more vastly critical of this version, especially regarding level of detail with a mean 3.29 (0.49), similarly lower ratings were reported for clarity [4.29 (0.76)], 4.43 (0.53) for language, and 3.71 (0.49) for content quality (**Table**

2).

Finally, for evaluations aimed at social media texts, all evaluators rated these texts highly. General public reviewers rated this version at 5.00 (0.00) for clarity, 4.75 (0.50) for detail, 4.50 (1.00) for language suitability, and 4.50 (1.00) for overall quality. Students and health professionals similarly gave high scores, 5.00 (0.00) for clarity, 4.40 (0.89) for detail, 4.80 (0.45) for language suitability, and 4.60 (0.55) for content quality. Communication professionals assigned perfect means (5.00 [0.00]) across all four criteria. Authors again provided more conservative ratings, with scores of 4.64 (0.50) for clarity, 4.09 (0.83) for detail, 4.73 (0.65) for language suitability, and 4.55 (0.52) for overall quality (**Table 2**).

Taken together, this evaluation of audience-adapted texts revealed consistently high mean scores across all four assessed criteria, regardless of the target audience. Notably, texts adapted for children and for the general population received particularly high ratings across all dimensions, with most mean scores approaching the maximum value of 5. Health students/professionals, communication professionals, and the general adult population tended to assign higher scores overall, while health managers and article authors demonstrated greater variability in their assessments, especially for Detailing of texts targeting health managers or social media platforms (**Figure 2**). Among all groups, article authors exhibited the greatest variability in their evaluations. These patterns suggest a general acceptability of the materials, with subtle differences in perceived quality depending on the evaluator profile (**Figure 3**).

In addition to quantitative ratings, anonymous evaluators were invited to provide open-ended comments on the chatbot-generated materials. A total of 68 comments were collected, distributed across the evaluated target audiences: “For children” (n = 14), “For general public” (n = 27), “For health managers” (n = 11), and “For social media” (n = 16) (**Table 3**). Across all categories, the overall tone of the comments was positive and constructive. For children’s materials, praise regarding Language was the most frequent (57.1%), while 42.9% of comments did not address Language. Information praise (28.6%) and Information criticism (28.6%) were also observed. Only 7.1% of comments suggested the material did not fully comply with its intended proposal. For health manager materials, comments were more evenly distributed: Information criticism (36.3%), Language praise (36.3%), and Proposal compliance (81.8%) were most common, though 18.2% indicated non-compliance with the proposal. For social media materials, the highest frequency of feedback related to Information criticism (37.5%) and Structure criticism (25.0%), reflecting the platform-specific communication challenges. A majority (81.3%) of comments judged the materials as compliant with the intended proposal. For general public materials, Language praise was dominant (66.7%), with a smaller proportion of comments addressing Information (only 29.6% provided any Information-related feedback). Most comments (85.2%) indicated that the outputs complied with the intended proposal. Across all targets, the high proportion of Proposal compliance comments indicates strong overall alignment between the outputs and their target audiences. A full compilation of the anonymous comments is provided in the **Supplementary Material**.

Comparative accuracy and context stability evaluation

A detailed comparison between *Maria Ciência* and the base large language model (ChatGPT 4o) was conducted across four conversational criteria: establishment of context; continuity of conversation; resilience to interruption; and return to context, stratified by target audience (**Table 4**). For *Maria Ciência*, mean scores were consistently high across all target audiences and criteria. For the general public, *Maria Ciência* achieved perfect continuity of conversation (5.00 ± 0.00) and high stability across all other dimensions (establishment of context: 4.75 ± 0.50; interruption of context: 4.75 ± 0.50; return to context: 4.75 ± 0.50). In comparison, the base GPT model showed slightly lower performance. For children, *Maria Ciência* again demonstrated superior performance, with perfect scores for interruption of context (5.00 ± 0.00) and strong scores across other criteria

(establishment of context: 4.66 ± 0.58; continuity: 4.66 ± 0.58; return to context: 4.66 ± 0.58). The base GPT model showed substantially lower performance in this category, particularly in continuity (3.00 ± 0.82) and establishment of context (3.25 ± 0.50), indicating challenges in maintaining audience-appropriate conversation flow for younger users.

For Social Media outputs, scores remained high, with means of 4.50 ± 0.58 for establishment of context and continuity of conversation, and 5.00 ± 0.00 for criterion resilience to interruption. In contrast, the base GPT model exhibited greater variability, particularly in outputs for Children and Health Managers. For Children, mean scores were notably lower in establishment of context (3.25 ± 0.96), continuity of conversation (3.00 ± 0.82), and resilience to interruption (3.00 ± 0.82), highlighting difficulties in maintaining and recovering conversational context. For Health Managers, while context establishment remained high (4.75 ± 0.50), performance dropped in criteria resilience to interruption (3.75 ± 0.96) and continuity of conversation (4.25 ± 0.50). Across all targets, the base GPT model showed more frequent context drift and reduced continuity compared to *Maria Ciência* (**Table 4**).

DISCUSSION

This study demonstrated that *Maria Ciência* is capable of producing audience-adapted science communication materials that are perceived as clear, accessible, and linguistically appropriate across a range of stakeholders. Materials tailored for children and the general public were especially well received, while outputs for health managers showed greater variability, reflecting the distinct informational demands of this audience. The more critical feedback from original article authors highlights the inherent challenge of balancing scientific precision with public accessibility.

Qualitative feedback from anonymous evaluators reinforced these trends, highlighting the clarity and perceived usefulness of the materials and offering constructive suggestions to further adapt tone and terminology for specific audiences. These results are encouraging, particularly in light of the urgent need to address health misinformation, which continues to erode public trust in science, hinder the implementation of health policies, and contribute to adverse health outcome [1,2]. The COVID-19 pandemic brought the urgency of this issue into sharp focus, with widespread infodemics interfering with disease prevention efforts and amplifying avoidable harm [2], emphasizing the importance of tools that support effective science communication.

By enabling the production of trusted, audience-specific materials, Maria Ciência can complement existing strategies for combating misinformation, which traditionally rely on reactive fact-checking or broad public health campaigns [22]. Unlike generic large language models, *Maria Ciência* demonstrated superior conversational stability and contextual accuracy, particularly for sensitive audiences such as children, critical for fostering health literacy from an early age. urthermore, previous studies have shown that GPT-based assistants specifically trained or configured with domain-relevant and ethically curated materials achieve higher performance and contribute to greater user trust and acceptance of the platform in public health and educational contexts [12,13]. The approach adopted by *Maria Ciência*, combining prompt engineering with thematic supervision, is consistent with these findings and reinforces the value of domain-specific configuration for science communication tools.

The development of *Maria Ciência* also reflects a reorientation of the role of researchers as not only producers of knowledge but as communicators and collaborators in public dialogue. By facilitating the creation of audience-specific materials, this tool can help bridge the gap between scientific knowledge and public understanding. Furthermore, its thematic supervision and alignment with inclusive language guidelines position it as a valuable ally in promoting equity in health communication. At the same time, structural barriers within the scientific community continue to

limit direct engagement with the public. Despite growing recognition of science communication as a core responsibility, academic structures still prioritize publication in peer-reviewed journals over community outreach [7]. Researchers often lack time, institutional support, or training to translate their findings into accessible formats. The data presented here suggest that even minimal support from tools like *Maria Ciência* can enable more scientists to participate meaningfully in public engagement. By providing formats aligned with the needs of different audiences and reducing the technical barriers to communication, the tool facilitates a shift toward more inclusive and democratic scientific practice.

As AI-driven science communication tools gain increasing relevance globally, it is important to consider their potential for application across diverse international contexts. Although initially developed and evaluated in Brazilian Portuguese, the architecture and prompting framework of *Maria Ciência*, being built on the GPT platform, allow the tool to operate in over 50 languages with the same ethical and technical standards[21]. This provides immediate potential for multilingual deployment in global health communication efforts. However, while the tool can linguistically adapt to multiple languages, effective application across different regions also requires attention to cultural nuances, health literacy variations, and sociolinguistic differences. Refining terminology, narrative styles, and framing of public health messages to align with local communication norms remains essential to ensure that AI-assisted science communication can support health literacy and misinformation prevention within diverse global public health ecosystems.

Despite these strengths, the study has limitations. The number of evaluators per target audience was not standardized, and some subgroups, such as health managers, were underrepresented. Author evaluations, while informative, were based on subjective perceptions and not validated against formal scientific fidelity criteria. The evaluation also focused on perception and usability rather than long-term impacts on knowledge retention or behavioral outcomes. Furthermore, the tool was primarily tested in Brazilian Portuguese; while the underlying architecture is multilingual, additional research is needed to validate its performance in other languages and cultural contexts. Even with these limitations, the evaluation provides a strong foundation to support the validity and utility of the tool. The systematic and transparent development process, combined with broad stakeholder engagement, underscores the potential of *Maria Ciência* to contribute meaningfully to health literacy and misinformation prevention. Its ability to outperform a baseline GPT model in contextual stability and accuracy further validates the importance of domain-specific configuration for public health applications.

In summary, *Maria Ciência* offers a promising avenue for enhancing knowledge translation and addressing the communication challenges posed by health misinformation. As international organizations and national governments have emphasized, combating the impacts of infodemics requires more than reactive correction. It necessitates proactive investment in tools and strategies that produce accessible, trusted, and culturally relevant information. By supporting the generation of audience-adapted materials and facilitating researcher engagement in public dialogue, *Maria Ciência* contributes to this broader effort to advance scientific literacy and strengthen public health resilience.

FUNDING: The study was supported by the Intramural Research Program of the Fundação Oswaldo Cruz (B.B.A.). B.S-P. received a fellowship from the Fundação de Amparo à Pesquisa da Bahia (FAPESB), K.V-S., J.O., G.P.R.C.S, S.M., R.R.S., and M.A-P., received fellowships from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil. B.B.A. is a sênior fellow from the CNPq.

ACKNOWLEDGMENTS: We gratefully acknowledge Ms. Luisa Vinagre and Mr. Rodrigo Pontes for their valuable contributions in generating the chatbot outputs used in this study. We also extend

our sincere thanks to all anonymous respondents who participated in the evaluation forms, and to Ms. Elze Leite for her administrative support throughout the project.

Conceptualization, M.A-P., K.V-S., and B.B.A.; Data verification and curation, B.S-P. and K.V-S.; Investigation, M.A-P., B.S-P., K.V-S., S.M., J.O., R.S.R., M.A-P. and B.B.A.; Formal analysis, M.A-P., B.S-P., G.P.R.C.S. , J.O. and K.V-S.; Funding acquisition, G.P.R.C.S. and B.B.A.; Methodology, M.A-P, KV-S., S.M., J.O., and G.P.R.C.S; Project administration, M.A-P.; Resources, M.A-P. and B.B.A.; Software, M.A-P., B.S-P., G.P.R.C.S. , J.O. and K.V-S.; Supervision, M.A-P. and B.B.A.; Writing—original draft, M.A-P., K.V-S., G.P.R.C.S., and B.S-P; and Writing—review and editing, all authors.

DATA SHARING STATEMENT: The data that support the findings of this study will be available upon reasonable request to the corresponding author of the study.

CONFLICT OF INTEREST: The authors have no conflicts of interest.

ETHICS STATEMENT: This study was conducted as a public opinion survey evaluating AI-generated science communication materials. All data were collected through voluntary and anonymous participation, without the collection of identifiable or sensitive personal information, and without any form of intervention or risk to participants. In line with international ethical standards and in accordance with Brazilian Resolution CNS 510/2016, which exempts public opinion research involving non-identified participants from requiring formal ethics committee approval, this study did not require submission to a research ethics committee.

Abbreviations
JMIR: Journal of Medical Internet Research
RCT: randomized controlled trial

REFERENCES

1. Oliveira TM de. Como enfrentar a desinformação científica? Desafios sociais, políticos e jurídicos intensificados no contexto da pandemia. Liinc em Revista Instituto Brasileiro de Informação em Ciência e Tecnologia; 2020 Dec 23;16(2):e5374–e5374. doi: 10.18617/liinc.v16i2.5374

2. Ferreira Caceres MM, Sosa JP, Lawrence JA, Sestacovschi C, Tidd-Johnson A, Rasool MHU, Gadamidi VK, Ozair S, Pandav K, Cuevas-Lou C, Parrish M, Rodriguez I, Fernandez JP. The impact of misinformation on the COVID-19 pandemic. AIMS Public Health 2022 Jan 12;9(2):262–277. PMID:35634019

3. Borges do Nascimento IJ, Pizarro AB, Almeida JM, Azzopardi-Muscat N, Gonçalves MA, Björklund M, Novillo-Ortiz D. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ 2022 Sep 1;100(9):544–561. PMID:36062247

4. BRASIL. Combate à desinformação na área da saúde: uma luta de todos. Ministério da Saúde. 2024. Available from: https://www.gov.br/saude/pt-br/assuntos/saude-com-ciencia/noticias/2024/maio/combate-a-desinformacao-na-area-da-saude-uma-luta-de-todos [accessed Sep 1,

2024]

5. Santos MLR, Paim MC, Soares CLM, Santos DM, Sande RS, Santos GR de M. Ações governamentais para enfrentamento da crise de desinformação durante a pandemia da Covid-19. Saúde debate Centro Brasileiro de Estudos de Saúde; 2022 Jun 24;45:187–204. doi: 10.1590/0103-11042021E213

6. Benecke O, DeYoung SE. Anti-Vaccine Decision-Making and Measles Resurgence in the United States. Global Pediatric Health SAGE Publications Inc; 2019 Jan 1;6:2333794X19862949. doi: 10.1177/2333794X19862949

7. Araújo-Pereira M, Andrade BB. The heroic journey of young Brazilian scientists: challenges and opportunities. The Lancet Regional Health – Americas Elsevier; 2024 May 1;33. doi: 10.1016/j.lana.2024.100735

8. Brownell SE, Price JV, Steinman L. Science Communication to the General Public: Why We Need to Teach Undergraduate and Graduate Students this Skill as Part of Their Formal Scientific Training. J Undergrad Neurosci Educ 2013 Oct 15;12(1):E6–E10. PMID:24319399

9. Hunter P. The communications gap between scientists and public. EMBO reports John Wiley & Sons, Ltd; 2016 Nov;17(11):1513–1515. doi: 10.15252/embr.201643379

10. Garrido-Merchán EC, Arroyo-Barrigüete JL, Borrás-Pala F, Escobar-Torres L, Ibarreta CM de, Ortiz-Lozano JM, Rua-Vieites A. Real Customization or Just Marketing: Are Customized Versions of Chat GPT Useful? arXiv; 2023. doi: 10.48550/arXiv.2312.03728

11. OpenAI. Introducing GPT-4.5. 2025. Available from: https://openai.com/index/introducing-gpt-4-5/ [accessed Jun 2, 2025]

12. Hakim VGA, Paiman NA, Rahman MHS. Genie-on-demand: A custom AI chatbot for enhancing learning performance, self-efficacy, and technology acceptance in occupational health and safety for engineering education. Computer Applications in Engineering Education 2024;32(6):e22800. doi: 10.1002/cae.22800

13. Pu J, Hong J, Yu Q, Yu P, Tian J, He Y, Huang H, Yuan Q, Tao L, Peng Z. Accuracy, satisfaction, and impact of custom GPT in acquiring clinical knowledge: Potential for AI-assisted medical education. Med Teach 2025 Feb 2;1–7. PMID:39893644

14. Cunda BV. Na ponta das línguas : pequeno glossário para apoiar o enfrentamento do estigma e da discriminação. 2023; Available from: https://lume.ufrgs.br/handle/10183/267984 [accessed May 3, 2025]

15. UNAIDS. UNAIDS Terminology Guidelines. 2024. Available from: https://www.unaids.org/en/resources/documents/2024/terminology_guidelines [accessed May 3, 2025]

16. UNICEF. Terminology guidelines | UNICEF. 2016. Available from: https://www.unicef.org/documents/terminology-guidelines [accessed May 3, 2025]

17. Frontiers for Young Minds: Science for kids, edited by kids. Available from: https://kids.frontiersin.org/ [accessed May 3, 2025]

18. Antonof GS, Rodrigues M da S, Mateus Schmitd. Diálogos: um guia para divulgação científica

em linguagem simples. Universidade de Brasilia; 2023. Available from: https://bdm.unb.br/bitstream/10483/38689/2/2023_GabrielAntonoff_MateusRodrigues_MateusSilva_produto.pdf

19. Paraguassu LB, Costa VM da. Guia de linguagem simples do ICICT. Fundação Oswaldo Cruz. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde.; 2023; Available from: https://www.arca.fiocruz.br/handle/icict/61432 [accessed May 3, 2025]

20. Profissão Biotec. Home - Profissão Biotec. 2025. Available from: https://profissaobiotec.com.br/ [accessed May 3, 2025]

21. OpenAI. How to change your language setting in ChatGPT | OpenAI Help Center. 2025. Available from: https://help.openai.com/en/articles/8357869-how-to-change-your-language-setting-in-chatgpt [accessed Jun 5, 2025]

22. Tiwary A, Rimal A, Paudyal B, Sigdel KR, Basnyat B. Poor communication by health care professionals may lead to life-threatening complications: examples from two case reports. Wellcome Open Res 2019 Jan 22;4:7. PMID:31448336

TABLES AND FIGURES

Table 1. Distribution of Reviewer Identities and Rating Frequencies by Target Audience.

Variable	For children N = 27	For health managers N = 35	For social media N = 24	For general public N = 35	p-value
Participant Classification, mean (%)					**<0.001**
Article author	10 (37.0)	7 (20.0)	11 (45.8)	5 (14.3)	
Communication professional	0 (0.0)	0 (0.0)	4 (16.7)	0 (0.0)	
General population (18+)	6 (22.2)	0 (0.0)	4 (16.7)	16 (45.7)	
Health manager	0 (0.0)	4 (11.4)	0 (0.0)	0 (0.0)	
Health student/professional	11 (40.7)	24 (68.6)	5 (20.8)	14 (40.0)	
Rating					
Clarity, mean (%)					0.205
1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
2	0 (0.0)	1 (2.9)	0 (0.0)	0 (0.0)	
3	0 (0.0)	2 (5.7)	0 (0.0)	0 (0.0)	
4	4 (14.8)	6 (17.1)	4 (16.7)	1 (2.9)	
5	23 (85.2)	26 (74.3)	20 (83.3)	34 (97.1)	
Detailing, mean (%)					0.275
1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
2	0 (0.0)	1 (2.9)	0 (0.0)	1 (2.9)	
3	2 (7.4)	8 (22.9)	4 (16.7)	1 (2.9)	
4	6 (22.2)	7 (20.0)	6 (25.0)	5 (14.3)	
5	19 (70.4)	19 (54.3)	14 (58.3)	28 (80.0)	
Language Adequacy, mean (%)					0.201
1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
2	0 (0.0)	1 (2.9)	0 (0.0)	0 (0.0)	
3	3 (11.1)	1 (2.9)	2 (8.3)	1 (2.9)	
4	4 (14.8)	12 (34.3)	2 (8.3)	6 (17.1)	
5	20 (74.1)	21 (60.0)	20 (83.3)	28 (80.0)	
Content Quality					0.114
1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
2	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
3	0 (0.0)	4 (11.4)	1 (4.2)	0 (0.0)	
4	7 (25.9)	12 (34.3)	7 (29.2)	7 (20.0)	
5	20 (74.1)	19 (54.3)	16 (66.7)	28 (80.0)	

Table Note: This table displays the frequency (n) and percentage (%) of reviewers' self-reported identity classifications and rating scores across four different domains, stratified by the number of reviews attributed to each target audience texts. P-values were calculated using fisher's exact test (for nominal categorical counts) or Kruskal–Wallis tests (for ordinal rating distributions), comparing across the four target-audience groups. A p-value < 0.05 denotes a statistically significant difference among audiences.

Table 2. Average (SD) evaluation scores by audience and stakeholder group for each adapted output

Target Audience	Evaluator	Clarity	Detailing	Language Suitability	Content Quality
For general public, mean (SD):	Authors	5.00 (0.00)	4.80 (0.45)	4.80 (0.45)	4.60 (0.55)
	General Public	4.94 (0.25)	4.56 (0.89)	4.62 (0.62)	4.75 (0.45)
	Students / Health Professionals	5.00 (0.00)	4.86 (0.36)	4.93 (0.27)	4.93 (0.27)
For children, mean (SD):	Authors	4.80 (0.42)	4.50 (0.71)	4.40 (0.84)	4.70 (0.48)
	General Public	5.00 (0.00)	4.83 (0.41)	4.83 (0.41)	4.83 (0.41)
	Students / Health Professionals	4.82 (0.40)	4.64 (0.67)	4.73 (0.65)	4.73 (0.47)
For health managers, mean (SD):	Authors	4.29 (0.76)	3.29 (0.49)	4.43 (0.53)	3.71 (0.49)
	Students / Health Professionals	4.83 (0.48)	4.58 (0.72)	4.62 (0.71)	4.67 (0.56)
	Health Managers	4.00 (1.41)	4.00 (1.41)	4.00 (0.82)	4.25 (0.96)
For social media, mean (SD):	Authors	4.64 (0.50)	4.09 (0.83)	4.73 (0.65)	4.55 (0.52)
	General Public	5.00 (0.00)	4.75 (0.50)	4.50 (1.00)	4.50 (1.00)
	Students / Health Professionals	5.00 (0.00)	4.40 (0.89)	4.80 (0.45)	4.60 (0.55)
	Communication Professionals	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)

Table Note: This table presents the central tendency and dispersion (mean [standard deviation, SD]) of four evaluation domains: A. Clarity, B. Detailing, C. Language Suitability, and D. Content Quality, across four target-audience contexts (“For children,” “For health managers,” “For social media,” “For general public”). Within each context, scores are shown separately for each evaluator subgroup (Authors; General public; Students/Health professionals; Health managers; Communication professionals). All ratings were provided on a five-point scale ranging from 1-5.

Table 3. Distribution of type of comments by Target Audience.

Category	For children N=14	For health managers N= 11	For social media N= 16	For general public N=27
Language, n (%):				
Criticism	0 (0)	1 (9.1)	2 (12.5)	2 (7.4)
Praise	8 (57.1)	4 (36.3)	5 (31.2)	18 (66.7)
Not applicable	6 (42.9)	6 (54.6)	9 (56.3)	7 (25.9)
Information, n (%):				
Criticism	4 (28.6)	4 (36.3)	6 (37.5)	5 (18.5)
Praise	4 (28.6)	2 (18.2)	1 (6.2)	3 (11.1)
Not applicable	6 (42.9)	5 (45.5)	9 (56.3)	19 (70.4)
Structure, n (%):				
Criticism	3 (21.4)	2 (18.2)	4 (25.0)	2 (7.4)
Praise	3 (21.4)	0 (0)	0 (0)	0 (0)
Not applicable	8 (57.1)	9 (81.8)	11 (68.8)	25 (92.6)
Proposal, n (%):				
Does not comply with proposal	1 (7.1)	2 (18.2)	3 (18.7)	4 (14.8)
Complies with proposal	13 (92.9)	9 (81.8)	13 (81.3)	23 (85.2)

Table Note: This table displays the frequency (n) and percentage (%). Comments were categorized into four domains: Language, Information, Structure, and Proposal. For each domain, comments were classified as Criticism, Praise, or Not applicable (for Language, Information, and Structure), or as Complies with proposal / Does not comply with proposal (for Proposal). Comments could be assigned to multiple domains and categories.

Table 4. Average (SD) evaluation scores of Comparative accuracy and context stability evaluation

Target Audience	Evaluator	Establishment of a context	Continuity of conversation	Interruption of context	Return to context
For general public, mean (SD):	*Maria Ciência*	4.75 (0.50)	5.00 (0.00)	4.75 (0.50)	4.75 (0.50)
	GPT base	4.25 (0.96)	4.25 (0.50)	4.25 (0.50)	4.75 (0.50)
For children, mean (SD):	*Maria Ciência*	4.66 (0.58)	4.66 (0.58)	5.00 (0.00)	4.66 (0.58)
	GPT base	3.25 (0.50)	3.00 (0.82)	3.75 (0.50)	4.50 (0.60)
For health managers, mean (SD):	*Maria Ciência*	4.75 (0.50)	5.00 (0.00)	4.50 (0.58)	4.50 (0.58)
	GPT base	4.50 (0.58)	4.50 (1.00)	4.25 (0.50)	5.00 (0.00)
For social media, mean (SD):	*Maria Ciência*	4.50 (0.58)	4.75 (0.50)	5.00 (0.00)	4.00 (0.00)
	GPT base	4.50 (0.58)	4.50 (0.58)	4.50 (0.58)	5.00 (0.00)

Table note: This table presents the central tendency and dispersion (mean [standard deviation, SD]) of four accuracy metrics: establishment of a context (A), continuity of a conversation (B), interruption of context (C) and return to context (D), across four target-audience contexts (“For children,” “For health managers,” “For social media,” “For general public”). Within each context, scores are shown separately for each model (“Maria Ciência” and “GPT base”). All ratings were provided on a five-point scale ranging from 1-5.

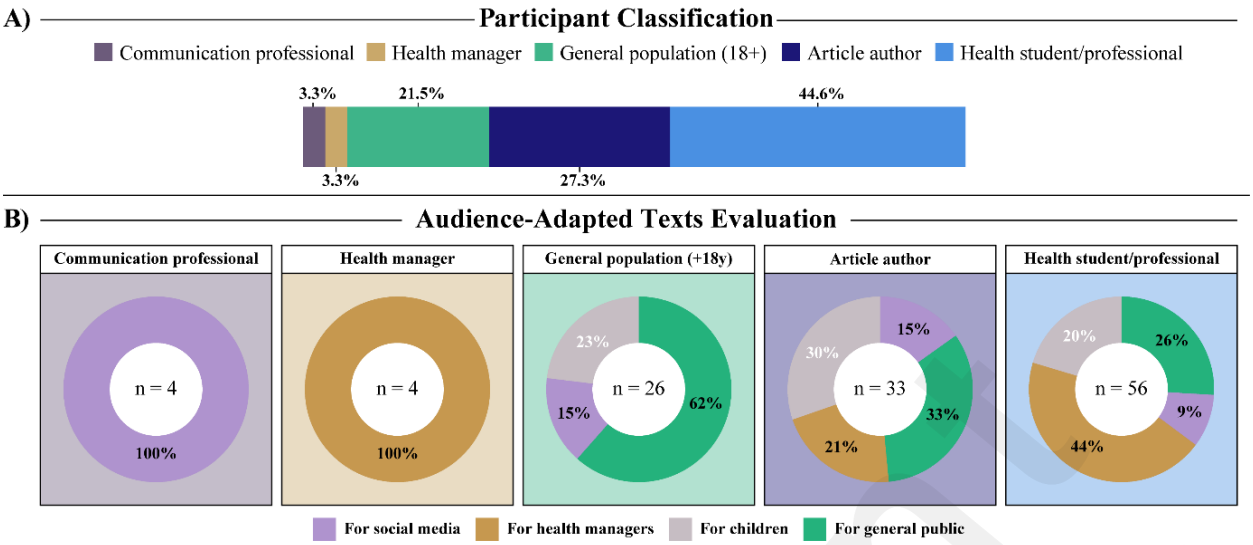


Figure 1. Participant Identities and Audience-Specific Text Evaluation Preferences. Figure represents: A) Distribution of study participants across specific self-provided identity classifications. Each bar segment represents the proportion of total respondents within each category, including Communication professionals, Health managers, General population (18+), Article authors, and Health students/professionals. B) Donut charts summarizing the types of audience-adapted texts proportionally reviewed by each participant identity classification. Segment colors represent the percentage of texts tailored for different audiences, including: i) For social media (lavender), ii) For health managers (golden), iii) For children (gray), and iv) For the general public (green). Each donut reflects the distribution of text types reviewed by participants within a specific identity group.

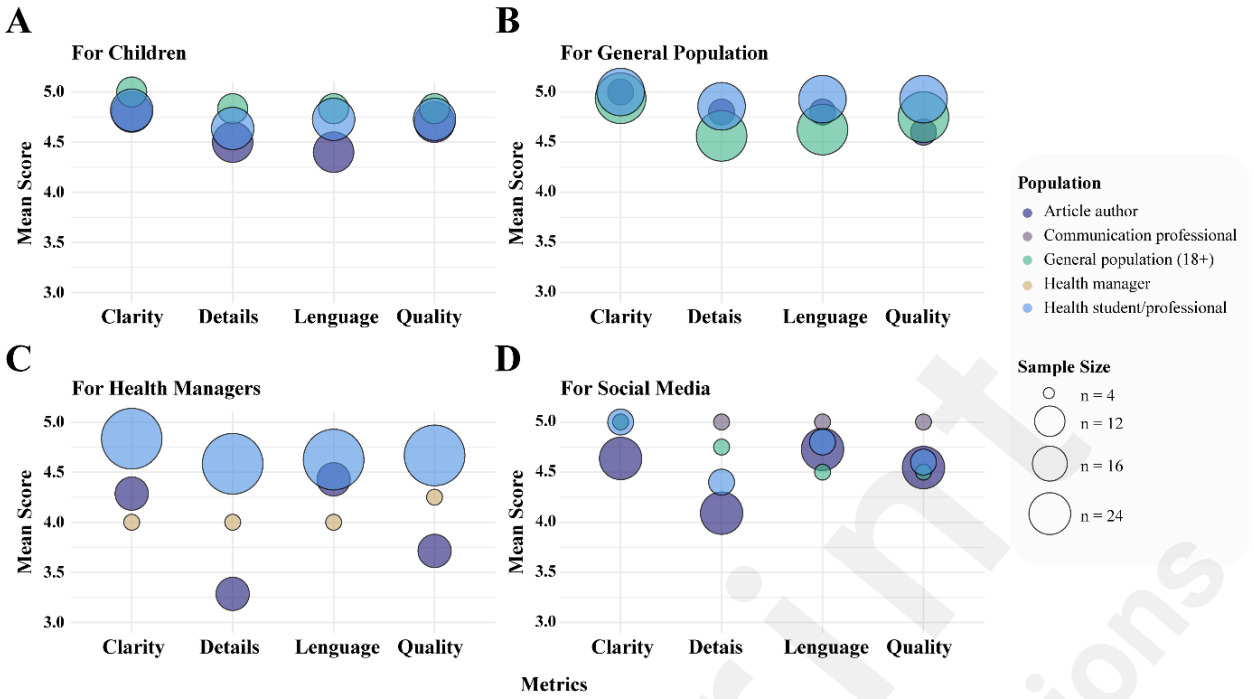


Figure 2. Mean Evaluation Scores of Audience-Adapted Texts by Criterion and Evaluator Population. Bubble plot presenting the mean scores for each evaluation criterion, Clarity, Detailing, Language Adequacy, and Content Quality, across four types of audience-adapted texts: A) For children; B) For General Population; C) For health managers; and D) For Social Media. Each colored bubble represents a different population of evaluators, as indicated in the legend on the right. The size of each bubble is proportional to the number of respondents from that population who rated the corresponding question for each text type.

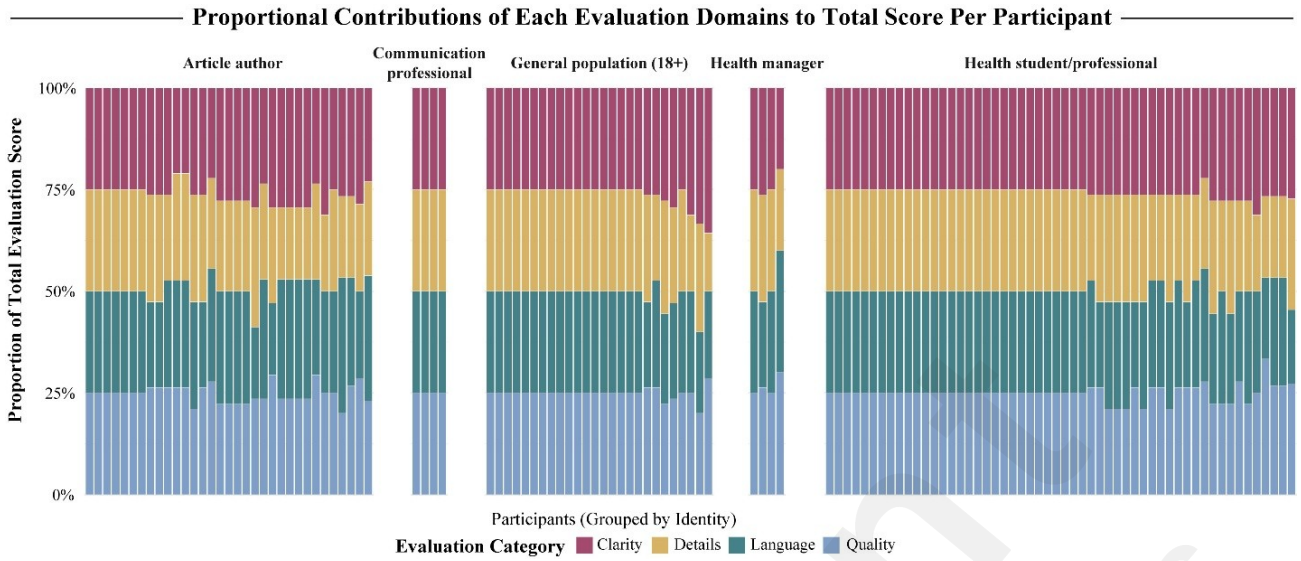
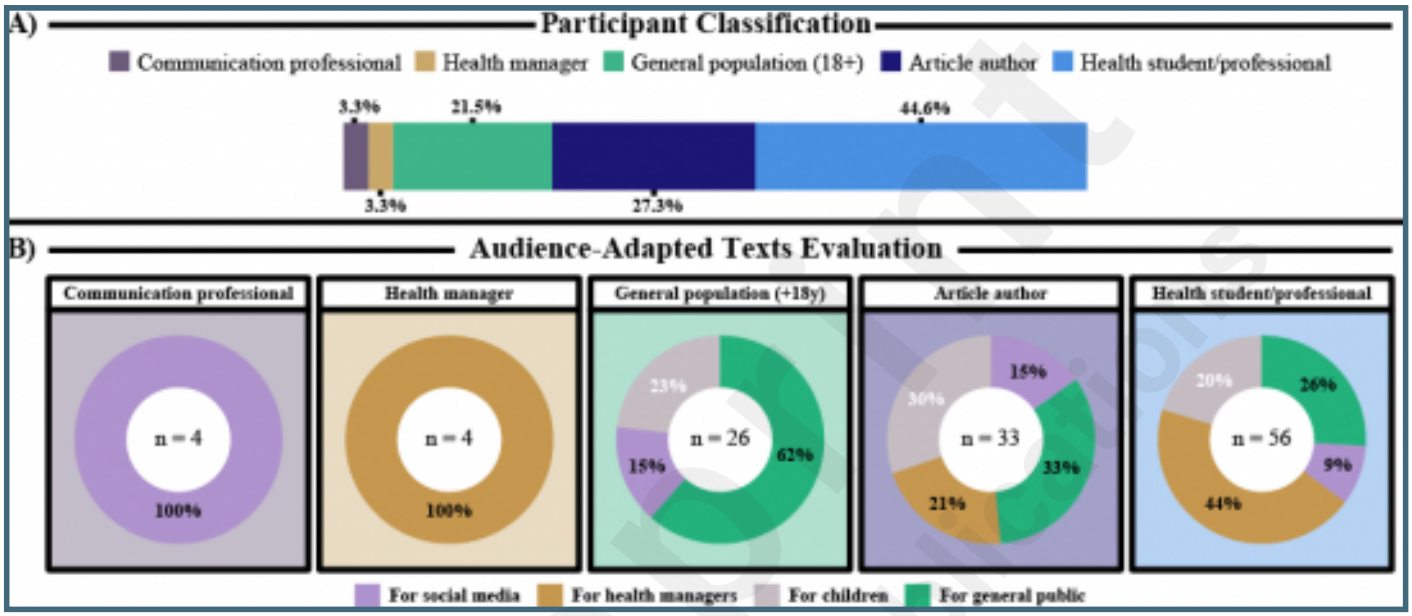


Figure 3. Proportional Contributions of Evaluation Domains to Each Participant's Total Score, Grouped by Reviewer Identity. 100% Stacked Bar Chart in which each vertical bar represents an individual participant's normalized total evaluation score (scaled to 100%), with participants organized along the x-axis and grouped by self-reported identity: *Article author*, *Communication professional*, *General population (18+)*, *Health manager*, and *Health student/professional*. Within each bar, colored segments depict the relative weight of each of the four evaluation categories, Quality (light blue), Language (teal), Detailing (gold), and Clarity (maroon), in that participant's overall total rating (sum of individual 1-5 domain scores). The y-axis indicates the percentage contribution of each category to the participant's total score.

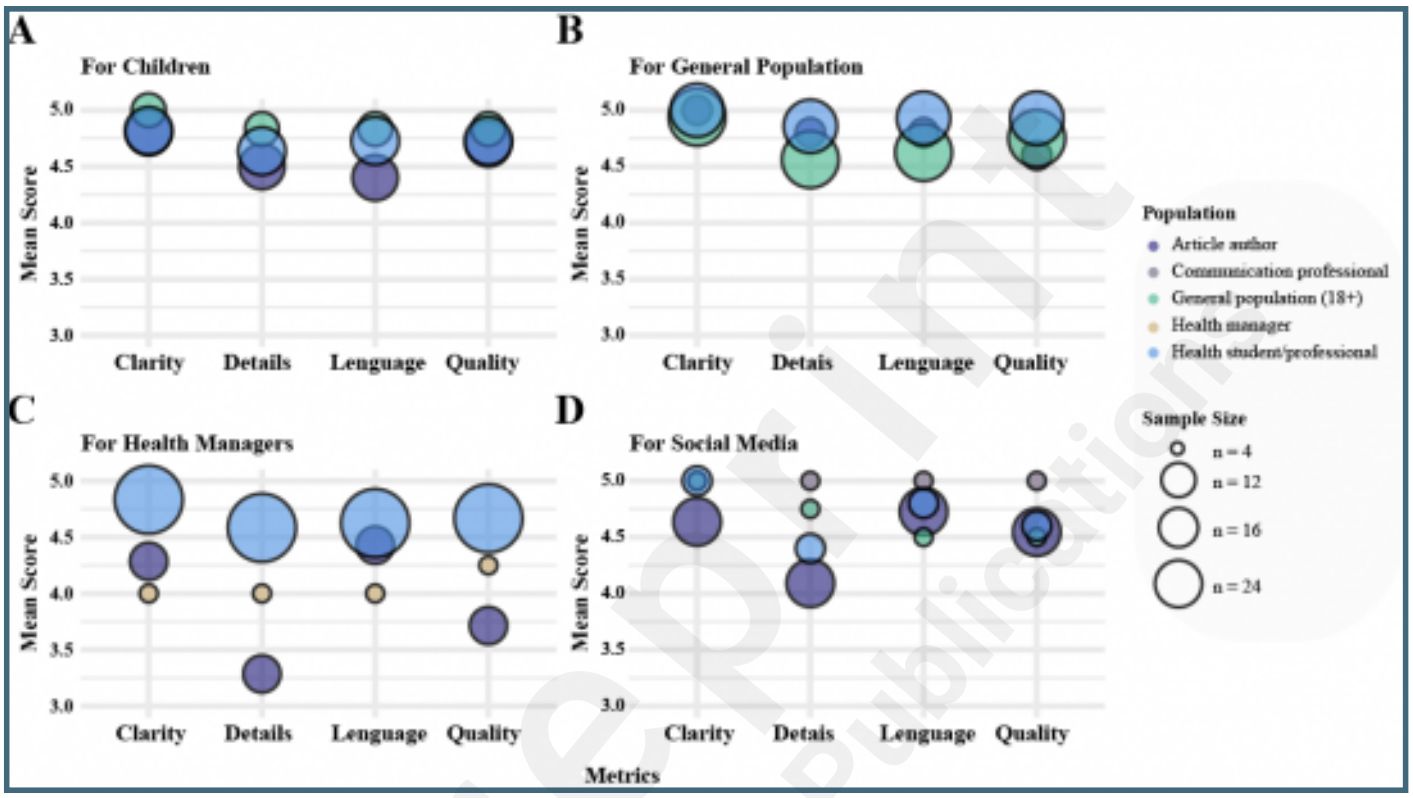
Supplementary Files

Figures

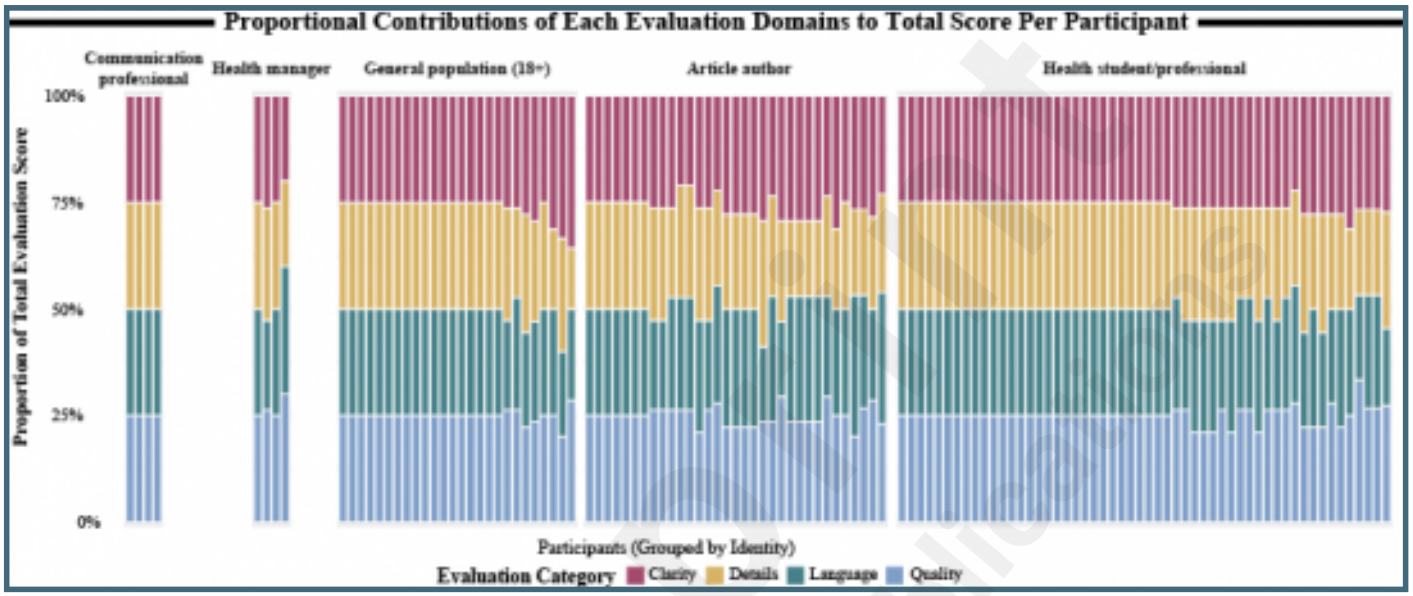
Participant Identities and Audience-Specific Text Evaluation Preferences. Figure represents: A) Distribution of study participants across specific self-provided identity classifications. Each bar segment represents the proportion of total respondents within each category, including Communication professionals, Health managers, General population (18+), Article authors, and Health students/professionals. B) Donut charts summarizing the types of audience-adapted texts proportionally reviewed by each participant identity classification. Segment colors represent the percentage of texts tailored for different audiences, including: i) For social media (lavender), ii) For health managers (golden), iii) For children (gray), and iv) For the general public (green). Each donut reflects the distribution of text types reviewed by participants within a specific identity group.



Mean Evaluation Scores of Audience-Adapted Texts by Criterion and Evaluator Population. Bubble plot presenting the mean scores for each evaluation criterion, Clarity, Detailing, Language Adequacy, and Content Quality, across four types of audience-adapted texts: A) For children; B) For General Population; C) For health managers; and D) For Social Media. Each colored bubble represents a different population of evaluators, as indicated in the legend on the right. The size of each bubble is proportional to the number of respondents from that population who rated the corresponding question for each text type.



Proportional Contributions of Evaluation Domains to Each Participant's Total Score, Grouped by Reviewer Identity. 100% Stacked Bar Chart in which each vertical bar represents an individual participant's normalized total evaluation score (scaled to 100%), with participants organized along the x-axis and grouped by self-reported identity: Article author, Communication professional, General population (18+), Health manager, and Health student/professional. Within each bar, colored segments depict the relative weight of each of the four evaluation categories, Quality (light blue), Language (teal), Detailing (gold), and Clarity (maroon), in that participant's overall total rating (sum of individual 1-5 domain scores). The y-axis indicates the percentage contribution of each category to the participant's total score.



Multimedia Appendixes

Untitled.
URL: http://asset.jmir.pub/assets/631edcdde1665877ee59057fd2a0f57c.docx