ABOUTOpen
Science Metrics

AboutOpen | 2024; 11: 77-88
ISSN 2465-2628 | DOI: 10.33393/ao.2024.3302

ORIGINAL RESEARCH ARTICLE

Open Access

Enhancing regulatory affairs in the market placing of new medical devices: how LLMs like ChatGPT may support and simplify processes

Fabio di Bello[1], **Eliana Russo**[2], **Matteo Sartori**[3]

[1]Senior Customer Training Manager and AI Expert, Wiley Global Technology CAST (Centro Studi Avanzati e Tecnologie) - Università di Chieti, Chieti - Italy
[2]Head of Regulatory Affairs, PRINEOS S.r.l., Milano - Italy
[3]Medical Expert of Regulatory Affairs, PRINEOS S.r.l., Milano - Italy

ABSTRACT

The market placing of a medical device in compliance with the requirements of EU Regulation 2017/745 (Medical Device Regulation) demands advanced regulatory expertise and a high level of detail and depth, inevitably leading to significant human and time resources. In an era where Artificial Intelligence (AI) is already present in various aspects of daily life, the potential and opportunity to use AI tools in the scientific field, such as the CE marking process of medical devices, are being explored. This process consists of several phases and related activities, some of which have been chosen as significant examples to evaluate how and to what extent AI can add value in achieving their compliance.

The article presents the overall results in terms of performance and reliability derived from generative AI tests using Large Language Models, such as ChatGPT, applied to some of the processes necessary for the market placing of a medical device. The method used focuses on the relationship between prompt quality and output quality, demonstrating the importance of prompt engineering in using these tools effectively alongside regulatory processes. It also emphasizes the need for end-users to have education, training, and understanding of the mechanisms of generative AI to optimize performance.

Keywords: Artificial Intelligence (AI), CE marking, ChatGPT, Large Language Models (LLMs), Medical Device Regulation (MDR), Prompt Engineering, Regulatory Affairs

Introduction

The medical device industry is facing increasing regulatory complexity following the implementation of Regulation (EU) 2017/745 (Medical Device Regulation, MDR), which requires strict regulatory compliance to demonstrate the efficacy and safety of medical devices that are to be made available on the market or are already on the market, including the CE marking process. This process needs significant human and time resources, not only for the drafting of detailed documentation and the execution of thorough clinical evaluations but also for conducting post-market surveillance and other management processes (1).

Simultaneously, the advent of Artificial Intelligence (AI) is revolutionizing various fields, offering opportunities for efficiency and innovation. Among AI technologies, Large Language Models (LLMs) like ChatGPT promise automation and simplification of complex processes.

This article explores the potential of LLMs in supporting regulatory affairs in the medical device field by critically evaluating their effectiveness, emphasizing prompt engineering, and highlighting the need for user training to ensure a performant approach to AI.

Current regulatory scenario

More than three years have passed since the European Regulation 2017/745 came into force. In the initial months, medical device stakeholders had little certainty and many doubts about how to comply with the new regulatory framework. Today, all agree that it requires increasingly stringent and impactful requirements to demonstrate and ensure the efficacy and safety of medical devices (1-3,6).

Primarily, manufacturers of medical devices were required to verify the state of the art of their existing products on the market to fill gaps resulting from incomplete technical and clinical data characterizing the device. Even more so for newly designed medical devices, the main activity was understanding how to apply the MDR to enable the CE marking of the product.

Received: September 13, 2024
Accepted: September 23, 2024
Published online: September 27, 2024

Corresponding author:
Fabio di Bello
email: fdibello@wiley.com

© 2024 The Authors. This article is published by AboutScience and licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).
Commercial use is not permitted and is subject to Publisher's permissions. Full information is available at www.aboutscience.eu

The requirements dictated by Regulation 2017/745 have been raised to such an extent that they sometimes necessitate a new classification of the device based on rules and risk classes, a new drafting of technical documentation to fully meet Annex II and Annex III of the MDR, and an adequacy check of the clinical evaluation of the device, often resulting in substantial revision due to a lack of performance, safety, and benefit-risk evidence and data. This includes risk analysis, usability of the device, and post-market surveillance process that must be actively conducted on the market by the manufacturer to verify whether the device achieves the intended use defined (1-7).

This significant effort, multiplied by the number of medical devices in the manufacturer's portfolio, has a strong impact in terms of time, resources, and consequently money, especially for small and medium-sized enterprises (SMEs) with limited investment capacity.

There is no doubt that the acquisition of skills, whether through internal resources or external qualified experts, is a source of optimization. Being competent, aware, and having high-level expertise improves timelines and makes processes, including the CE marking of a medical device, more linear and logical, thus shorter in reaching the set goal.

Certainly, other methods, tools, and resources can complement the manufacturer or stakeholders in the field to facilitate the market placing or maintenance of a medical device. New methods are under evaluation and study in terms of applicability.

AI and regulatory affairs: a synergistic approach

Faced with the complexity of MDR requirements, AI, particularly LLMs, can play a role in supporting the implementation of regulatory processes by enhancing their efficiency and compliance wherever possible (8).

The main activities required from the manufacturer for the market placing of a medical device can be identified as follows:

1. Drafting the technical documentation of the medical device: the generation and review of technical documentation can be time-consuming (1). ChatGPT could assist in drafting detailed descriptions of devices, risk management plans, general safety and performance requirements checklist, and technical reports (8).
2. Generating risk analysis of the device (1): LLMs could support manufacturers in identifying risks associated with the medical device and in pinpointing risk mitigation measures (8).
3. Drafting and approval of the clinical evaluation of the medical device (2,4,5): LLMs could aid in collecting clinical data from scientific literature to generate clinical evaluation reports. They could summarize findings and highlight key points (8).
4. Conducting usability verification of the medical device (7): ChatGPT could generate a usability test based on precise prompts that direct it to evaluate the specific use of the medical device (8).
5. Implementation of post-market surveillance processes (1,6): LLMs could assist in creating survey questionnaires as a method for collecting efficacy and safety data on the medical device directly from market users (8).

Tests have been conducted to evaluate the use of AI tools (ChatGPT, LLMs) in some of the aforementioned regulatory activities to assess their ability to function as regulatory expertise, support it, or, conversely, where they may fail to provide satisfactory performance.

The LLMs: what ChatGPT is and how it works

ChatGPT is an advanced language model developed by OpenAI, capable of understanding and generating natural text, facilitating many activities traditionally performed by specialists. ChatGPT is based on the Generative Pre-trained Transformer (GPT), a class of language models that uses deep neural networks to understand and generate text (9). The power of GPT lies in its architecture, which leverages language transformations in geometry and vector calculation to process information efficiently and flexibly (Fig. 1).

Figure 1 illustrates the basis of GPT's functioning: the ability to model natural language into geometric representations. Each word or phrase is converted into vectors, that is, points in a multidimensional space that capture the meaning and relationships between terms. This process, known as embedding, allows the model to manipulate and analyze language mathematically (Fig. 2). For example, words with similar meanings tend to be close in vector space, while words with different meanings are further apart. This approach enables GPT to understand the context and nuances of human language, improving the precision of generated responses.

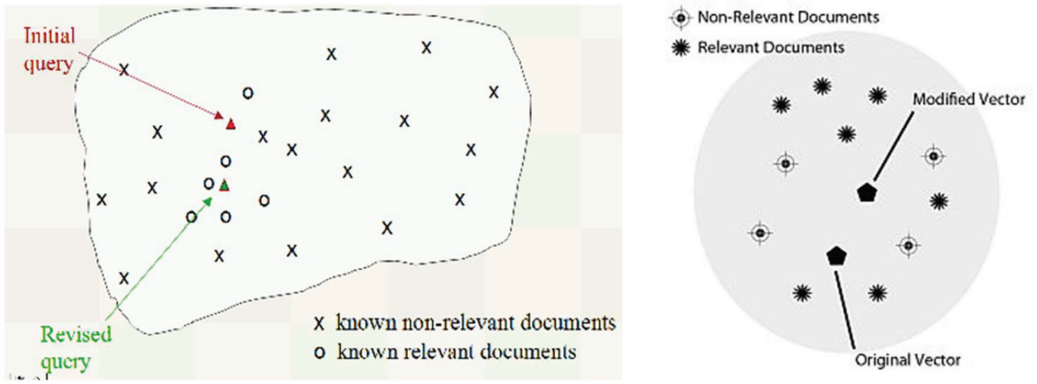
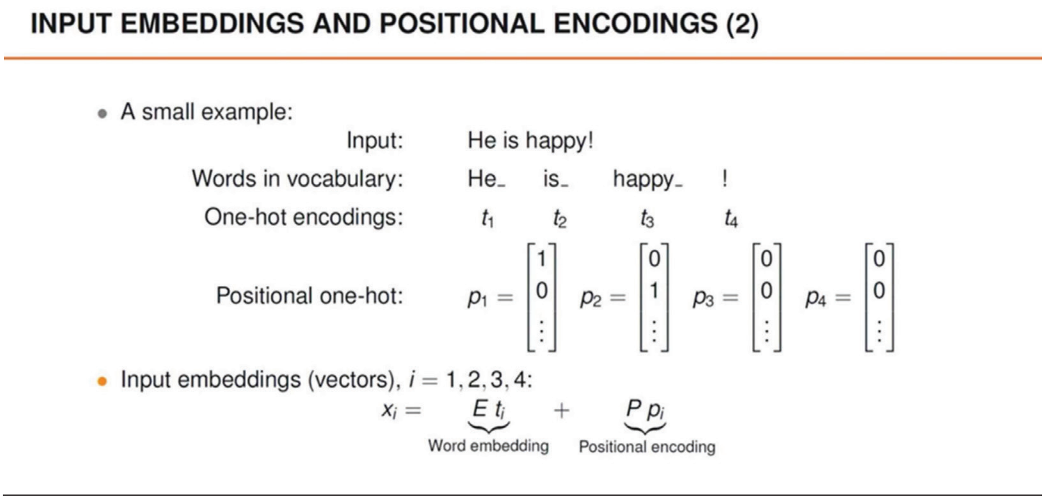


FIGURE 1 - Example of conversion of text into vectors in space.



INPUT EMBEDDINGS AND POSITIONAL ENCODINGS (2)

- A small example:

Input: He is happy!

Words in vocabulary: He_ is_ happy_ !

One-hot encodings: t_1 t_2 t_3 t_4

Positional one-hot: $p_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \end{bmatrix}$ $p_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \end{bmatrix}$ $p_3 = \begin{bmatrix} 0 \\ 0 \\ \vdots \end{bmatrix}$ $p_4 = \begin{bmatrix} 0 \\ 0 \\ \vdots \end{bmatrix}$

- Input embeddings (vectors), $i = 1, 2, 3, 4$:

$$x_i = \underbrace{E\,t_i}_{\text{Word embedding}} + \underbrace{P\,p_i}_{\text{Positional encoding}}$$

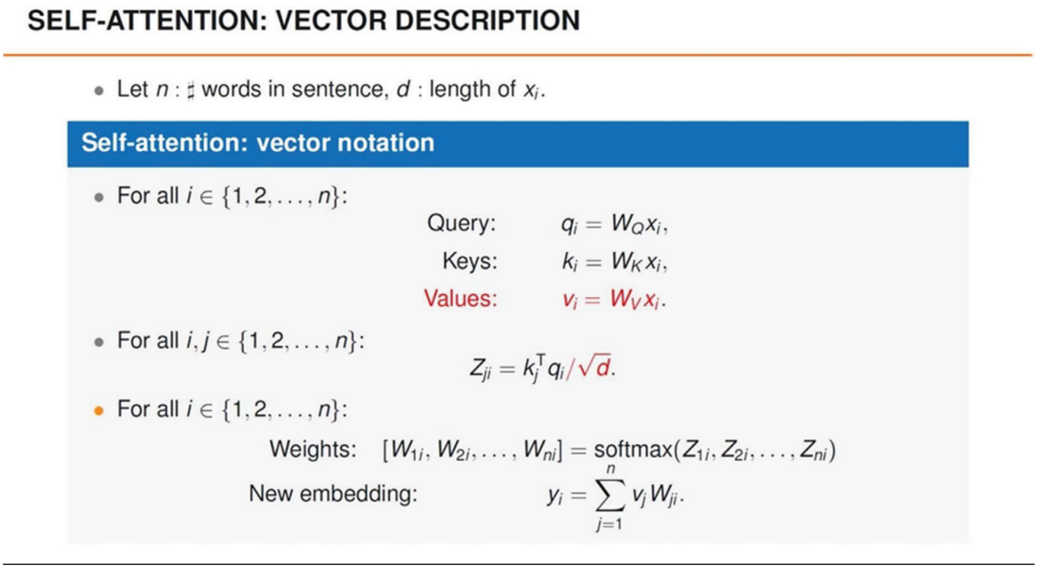
FIGURE 2 - Example of input embedding and positional encoding for a sentence.

The attention mechanism

A key component of GPT is the attention mechanism: it allows the model to focus on specific parts of the input text while generating a response. Attention helps identify and prioritize relevant information, enhancing the coherence and relevance of responses. The mechanism operates through a series of weights that determine the importance of each word relative to others. These weights are dynamically calculated during text processing, allowing GPT to adapt to different situations and linguistic contexts (Fig. 3). Figure 3 shows the SoftMax function, which underpins the mathematical encoding mechanism of attention.

As shown in Figure 3, Self-Attention is a fundamental mechanism in Transformer models that allows the model to weigh the importance of each word in a sequence relative to the others during natural language processing. Here is a brief explanation of the vectors involved:

1. **Input embedding:** each word in the sequence is represented by an embedding vector, which is a numerical representation of the word.
2. **Query (Q), Key (K), and Value (V) vectors:** for each word, three distinct vectors are created:
   - **Query (Q):** represents the word from which the "question" originates to assess its importance relative to the other words.
   - **Key (K):** describes the importance of a word relative to the others, acting as a "key" that will be compared with the query.
   - **Value (V):** contains the information that will be weighted and combined, representing the actual information to be extracted.
3. **Attention calculation:** to calculate how important a word is compared to the others, the dot product between the Query and Key vectors is computed. This value is then normalized (usually with a SoftMax function) to produce weights.



SELF-ATTENTION: VECTOR DESCRIPTION

- Let n : ♯ words in sentence, d : length of x_i.

Self-attention: vector notation

- For all $i \in \{1, 2, \ldots, n\}$:

Query: $q_i = W_Q x_i$,

Keys: $k_i = W_K x_i$,

Values: $v_i = W_V x_i$.

- For all $i, j \in \{1, 2, \ldots, n\}$:

$$Z_{ji} = k_j^{\mathrm{T}} q_i / \sqrt{d}.$$

- For all $i \in \{1, 2, \ldots, n\}$:

Weights: $[W_{1i}, W_{2i}, \ldots, W_{ni}] = \mathrm{softmax}(Z_{1i}, Z_{2i}, \ldots, Z_{ni})$

New embedding: $y_i = \sum_{j=1}^{n} v_j W_{ji}$.

FIGURE 3 - Self attention mechanism and Softmax function calculation.

© 2024 The Authors. Published by AboutScience - www.aboutscience.eu

4. **Application of weights:** the calculated weights are used to combine the Value vectors, giving more importance to the words considered more relevant according to the weights.
5. **Final output:** the result of the process is a new vector for each word, representing a weighted combination of the other words in the context.

The self-attention mechanism allows the model to capture complex and dynamic relationships between words in a sequence, enhancing context understanding compared to other traditional methods (9,10).

Differences from rule-based systems

Old natural language processing systems were based on predefined rules manually created by developers. These systems worked well for specific tasks but were rigid and difficult to adapt to new contexts. The rules needed continuous updating and expansion to handle new scenarios, requiring significant human effort.

In contrast, ChatGPT uses machine learning to extract knowledge from large amounts of textual data. This approach allows the model to generalize better and manage a variety of tasks without the need for continuous manual intervention (10). The flexibility and continuous learning capability of GPT make it possible to support many activities in regulatory affairs, ranging from document drafting to clinical data research and analysis.

Token consumption: optimization, limits, and methods

One of the crucial aspects of using ChatGPT is managing token consumption. Each request to GPT is processed based on a number of tokens, where a token can be a character, a word, or part of a word. GPT models have a maximum token limit they can handle in a single request, usually ranging from 2048 to 4096 tokens.

When processing large volumes of documents, the token limit can become a significant issue. If the text exceeds the token limit, responses may be partial or incomplete.

There are various methods to optimize token consumption and mitigate associated problems. Some of these include:

1. **Text segmentation:** Divide the document into smaller segments that fall within the token limit. Each segment can be processed separately and then combined to obtain the complete response.
2. **Optimized prompting:** Create more concise and specific prompts to reduce the number of tokens used in the request. This requires strategic formulation of questions to maximize efficiency.
3. **Iteration and synthesis:** Use partial responses to iteratively build a complete answer. This method involves requesting multiple successive responses and synthesizing the obtained information.
4. **Using more advanced models:** Adopt newer and more advanced versions of GPT models that may have a higher token capacity or better processing efficiency.

Methodology for evaluating ChatGPT

To assess the comprehensiveness and reliability of LLMs like ChatGPT in processes, it is essential to use a structured and rigorous methodology. This approach involves comparing the outputs generated by ChatGPT with those produced by industry experts in selected activities.

1. **Definition of objectives**
   Before starting the tests, it is crucial to clearly define the evaluation objectives:
   - Determine the accuracy of ChatGPT's responses.
   - Assess the comprehensiveness of the information provided.
   - Measure the coherence and relevance of the outputs against industry standards.
2. **Selection of activities to be performed**
   Identify the specific activities to be tested for the evaluation.
3. **Development of prompts**
   Create detailed and specific prompts to guide ChatGPT in generating outputs. Prompts should be formulated to obtain complete and relevant responses.
4. **Collection of outputs**
   Use the developed prompts to generate outputs from ChatGPT. Simultaneously, request a group of industry experts to produce outputs for the same activities. Ensure that the experts have extensive experience and knowledge in the field of interest.
5. **Evaluation criteria**
   Establish clear and objective criteria for evaluating the outputs. Suggested criteria include:
   - Accuracy: check if the information provided is correct and error-free.
   - Comprehensiveness: assess if the output covers all aspects required by the prompt.
   - Coherence: measure the logical coherence of the responses.
   - Relevance: verify if the information is pertinent and useful for the objective of the activity.
   - Clarity: ensure that the output is clear and understandable.
6. **Comparison and analysis**
   Compare the outputs generated by ChatGPT with those produced by experts using the established evaluation criteria. This comparison can be conducted in two phases:
   - Phase 1: Individual Evaluation
     Each output is independently evaluated by a team of reviewers.
   - Phase 2: Comparative Analysis
     Compare the evaluations to identify differences between ChatGPT and the experts. This includes identifying areas where ChatGPT excels and areas where it could improve, or conversely, where it may be lacking in comprehensiveness.
7. **Discussion of results**
   Analyze the comparison results to draw conclusions about the effectiveness of ChatGPT: discuss the strengths and

© 2024 The Authors. AboutOpen - ISSN 2465-2628 - www.aboutscience.eu/aboutopen

weaknesses identified, as well as possible implications for the future use of ChatGPT in the application field. Also, consider suggestions for improving prompts and optimizing token usage.

Training and optimization

Based on the results, develop a training program for users that includes:

- Best practices for creating effective prompts.
- Strategies to optimize token consumption.
- Practical examples and case studies to illustrate the use of ChatGPT in regulatory processes.

Prompt engineering and output quality

Prompt engineering, or the art of creating precise and detailed input requests for language models like ChatGPT, is a crucial element that directly impacts the quality of the generated outputs. Good prompt engineering can make the difference between a vague, irrelevant output and a precise, useful response. Below are the fundamental principles of prompt engineering and how it can be optimized to improve output quality.

Fundamental principles of prompt engineering

1. Clarity and specificity
   - Prompts must be clear and specific. A well-defined prompt helps the model understand exactly what is being requested. For example, instead of asking, “Describe the medical device,” a more specific prompt might be, “Describe the main features and clinical applications of the new medical device XYZ.”
2. Appropriate context
   - Providing the appropriate context is essential for obtaining relevant responses. A prompt that includes context helps the model in narrowing down the possible answers. For example, “Describe the safety features and emergency protocols for the medical device XYZ used in hospital settings.”
3. Structure and formatting
   - Good structure and formatting of the prompt can significantly improve the quality of responses. It is advisable to use bullet points, numbered lists, or separate paragraphs for different sections of the request. For example: “Generate a usability test report for the medical device XYZ that includes:
     1. Test Objective
     2. Methodology
     3. Key Results
     4. Recommendations”
4. Examples of responses
   - Providing examples of desired responses can help the model better understand the expected type of output. For example:
     “Create a questionnaire for the usability of a new medical device.” Here is an example of output:
     “The information on the packaging (device name, purpose, warnings) is:
     - Very clear
     - Clear
     - Somewhat clear
     - Not clear at all”

Constraints and guidelines

- Setting specific constraints and guidelines can improve the accuracy of responses. For example, if concise answers are required, specify this clearly in the prompt. “Provide a brief description (maximum 150 words) of the main features of the medical device XYZ.”

Optimization of prompts for token consumption

Optimizing token consumption is a critical aspect, especially when dealing with large volumes of documents. Here are some strategies to optimize prompts:

1. **Conciseness**
   - Avoid unnecessary words and keep the prompt as concise as possible without sacrificing clarity. For example, instead of saying, “Could you please provide me with a detailed description of...,” use “Describe in detail....”
2. **Focus**
   - Focus on key points and limit requests to a specific topic to reduce the number of tokens. For example, divide a complex prompt into multiple targeted chats.
3. **Removal of unnecessary information**
   - Eliminate redundant or unnecessary information that does not add value to the request. For example, avoid long introductions or preliminary explanations.

Example of prompt before optimization

“Hello, I need your help to create a detailed questionnaire that we can use to gather feedback from users about the new medical device we have recently developed. Could you include questions related to ease of use, perceived effectiveness, and any suggestions for improvements? Thank you so much!”

Example of prompt after optimization

“Create a questionnaire to gather feedback on the new medical device XYZ. Include questions on:
1. Ease of use
2. Perceived effectiveness
3. Suggestions for improvements”

Evaluation of output quality

To evaluate the quality of outputs generated by optimized prompts, it is important to use the following criteria:
1. Relevance: Is the output relevant to the request?
2. Completeness: Does the output cover all required aspects?

© 2024 The Authors. Published by AboutScience - www.aboutscience.eu

3. Clarity: Is the output clear and understandable?
4. Accuracy: Is the output correct and free of errors?
5. Comprehensiveness: Does the output provide an exhaustive answer?

AI applied to regulatory affairs: two examples based on MDR requirements

Survey

With the entry into force of Regulation 2017/745, manufacturers have been required to verify the availability and evaluate the quality of clinical evidence, and therefore clinical data, on their medical device to demonstrate its effectiveness and safety.

From an analysis of evidence, particularly in terms of quantity and quality, there is often a need to collect new clinical data on the device. It is known that the Regulation mandates clinical investigations for class III medical devices and implantable devices, except in some specific cases. The same requirement applies to innovative devices not yet on the market. Once these devices are marketed, as well as for legacy medical devices, that is, those already CE marked and thus commercially available, manufacturers can utilize post-market surveillance (PMS) processes to collect evidence and clinical data different from a clinical investigation, such as surveys. A survey is a questionnaire given to the medical device user, including patients, distributors, or healthcare professionals (doctors, pharmacists, etc.) who respectively use, distribute, or prescribe the medical device, to gather responses on the device's performance and safety (1,3,6).

A survey developed and distributed to collect clinical evidence on a substance-based medical device, specifically a gynecological cream with a barrier effect that helps restore the physiological pH of the vagina and allows the recovery of optimal conditions for the proliferation of natural lactobacilli, is one of the case studies used to apply AI to the regulatory requirement of PMS.

The survey, created by leveraging regulatory skills along with clinical and statistical expertise, begins by asking the medical device user about their age and the circumstances that led to the development of the vaginal condition. The medical device is used to alleviate signs and symptoms typically present in these circumstances or to prevent recurrent vaginal infections.

To gather data to evaluate the performance of the gynecological cream, the survey asks users to select the symptoms present, particularly the most persistent ones, assigning an intensity score from absent to very severe, both before and after using the medical device. This way, evidence of any improvements obtained with the therapy is collected.

Additionally, the survey asks whether using the device as a preventive treatment has led to a reduction in episodes of recurrent vaginal infections. To collect safety data on the device, the survey asks users if any undesirable effects or adverse reactions occurred after using the vaginal cream. If so, they are asked to specify which ones.

The survey concludes by asking for an overall judgment on the use of the medical device and whether the cream should be recommended to others with a similar vaginal condition. The data collected through the survey were analyzed and processed by statistical methodology experts to obtain objective data on the device's performance and safety, which are statistically significant.

After the design, execution, and finalization, including statistical analysis, of the survey that meets regulatory requirements, the case study continued with testing the generation using AI. The AI was given the prompt for a survey questionnaire to collect performance and safety data on the medical device, specifically a gynecological cream for restoring vaginal physiological conditions and preventing recurrent infections.

Prompt

"Write a complete survey that aims at collecting all necessary information regarding safety and performance for patients using a medical device such as a vaginal cream according to its intended use, that is, a gynecological cream with a barrier effect that helps restore the physiological pH of the vagina and allows the recovery of optimal conditions for the proliferation of natural lactobacilli. The survey needs to report an evaluation of the efficacy of the medical device in treating signs and symptoms before and after the application of the medical device."

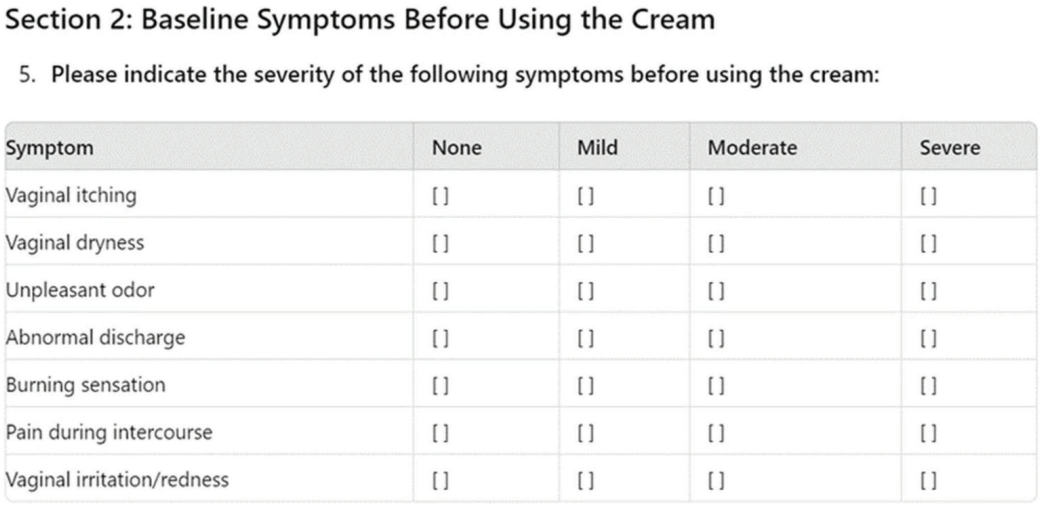
The survey questionnaire generated by ChatGPT begins by asking the user of the medical device about their age and general health conditions. Then, it requests information on which symptoms and their intensity the patient experienced before and after using the gynecological cream, as well as any manifestation of adverse events or new symptoms.

The survey concludes with a satisfaction judgment on the effectiveness of the medical device and whether it should be recommended to others. It is evident that AI was able to support the generation of a survey for the device of interest, covering most of the effectiveness and safety topics required for collecting clinical data on the medical device in the PMS process.

However, there is a lack of quantitative characterization of the severity level of the symptoms addressed by the cream, as ChatGPT provides, in this case, a descriptive-qualitative severity attribution of the symptoms based on a single prompt (Fig. 4), thus not allowing for statistical analysis of the collected data.

Regarding the safety questions, ChatGPT correctly suggests that the user should indicate any adverse events or new symptoms that may have occurred during or after treatment with the gynecological cream (Fig. 5).

Based on the material produced, the manufacturer of the medical device must review the survey suggested by AI through a detailed literary and statistical analysis to make the questionnaire suitable for generating the expected clinical performance and safety data. The AI output can, therefore, be given an overall rating of "Good" (Fig. 6), considering that in terms of relevance, it is pertinent to the request; in terms of completeness, it lacks some aspects as mentioned above, but it is clear, understandable, and sufficiently comprehensive.

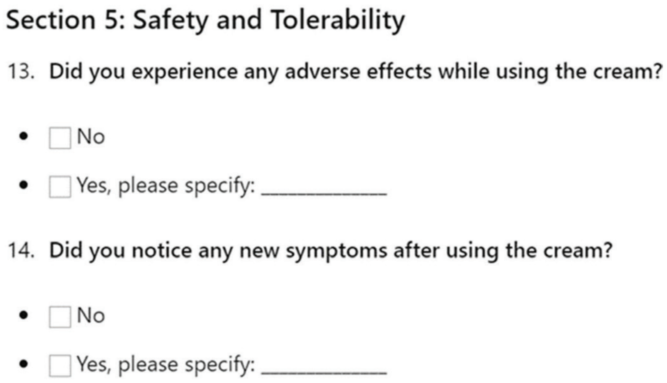


Section 2: Baseline Symptoms Before Using the Cream

5. **Please indicate the severity of the following symptoms before using the cream:**

Symptom	None	Mild	Moderate	Severe
Vaginal itching	[ ]	[ ]	[ ]	[ ]
Vaginal dryness	[ ]	[ ]	[ ]	[ ]
Unpleasant odor	[ ]	[ ]	[ ]	[ ]
Abnormal discharge	[ ]	[ ]	[ ]	[ ]
Burning sensation	[ ]	[ ]	[ ]	[ ]
Pain during intercourse	[ ]	[ ]	[ ]	[ ]
Vaginal irritation/redness	[ ]	[ ]	[ ]	[ ]

FIGURE 4 - Output of ChatGPT on the survey question regarding symptoms and their relative intensity before the use of the medical device.



Section 5: Safety and Tolerability

13. **Did you experience any adverse effects while using the cream?**

- ☐ No
- ☐ Yes, please specify: ____________

14. **Did you notice any new symptoms after using the cream?**

- ☐ No
- ☐ Yes, please specify: ____________

FIGURE 5 - Output of ChatGPT on the survey question regarding safety elements.

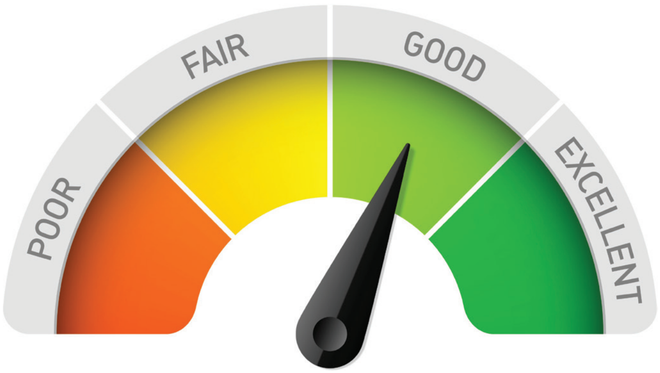


FIGURE 6 - Evaluation Dashboard of ChatGPT's output related to the case study on generating a survey for the medical device gynecological cream.

Usability

With the MDR, usability has become an essential requirement for obtaining CE marking of a medical device. The manufacturer must ensure that the device is effective and safe, and that the risks associated with its use are acceptable when compared to clinical benefits. The manufacturer must assess potential risks and errors that may occur during the intended use of the medical device and reasonably foreseeable misuse. For this reason, usability, which aims to protect the patient, has become an important element whose application allows for easy and appropriate use of the product in terms of user interface and confirmation of the effectiveness of labeling concerning readability and comprehension.

Based on the current state of the art, usability is governed by the IEC 62366 Standard "Application of usability engineering to medical devices," which states that the manufacturer's definition of usability engineering leads to the establishment of requirements that, if correctly implemented, can increase the likelihood that users will use the device without errors.

To evaluate, analyze, and improve the usability of the medical device, the manufacturer performs a usability test, which consists of a series of questions directed at the user about the ease of use of the product based on the provided instructions for use (IFU) (1,7).

A usability test developed and submitted to users of a medical device, such as a polyurethane breast implant, is one of the case studies used to apply AI to the regulatory usability requirement of a medical device.

The breast implant acts as a temporary replacement for removed tissue and supports the growth of autologous tissue in patients. The device, intended for use in adult female subjects, is used in patients who wish to undergo conservative and reconstructive breast surgery to restore the natural appearance of the breast.

The device is intended for use by qualified medical-surgical personnel who are provided with a usability questionnaire. The questionnaire starts by asking about the clarity of consulting the package labeling, as well as the ease of opening, and continues by investigating the convenience and ease of extraction from the packaging and hydration of the

© 2024 The Authors. Published by AboutScience - www.aboutscience.eu

device. It includes questions on the usability of the breast implant during surgery in terms of positioning, securing, and covering the device, as well as during potential removal or explant.

The questionnaire includes a section dedicated to collecting feedback on the clarity of the information provided by the manufacturer to the user regarding packaging and sterilization, materiovigilance, storage, disposal, warnings, precautions, adverse effects, residual risks, and complaints.

The questionnaire concludes with a request for an overall satisfaction rating on the clarity of the IFU and suggestions for their improvement.

Following the creation of the usability questionnaire that meets regulatory requirements, the case study continued with the generation using AI, which was given the prompt to develop a usability test for a breast implant.

Prompt (1st)

"I am attaching the file, which is the IFU of a medical device. Can you write a usability questionnaire, that is, a series of questions that allow us to evaluate the likelihood of human error or misuse of the device due to a lack of clarity in the IFU or difficulty in using the device according to the instructions provided by the manufacturer?"

To provide ChatGPT with adequate context to guide the generation process, the IFU document of the device was attached as a knowledge element to the prompt, requesting the creation of a usability questionnaire.

It is essential to share the IFU with the AI, considering that the objective of the usability test is to verify the clarity and comprehension of the instructions provided to the user of the specific device, in order to revise them if these lead to errors and, consequently, misuse during the test.

The usability questionnaire generated by ChatGPT remains generic without delving into the expected specifics of the medical device's use. It formulates questions on the clarity and completeness of the instructions (Fig. 7), as well as on the ease of use of the device, without asking the user to effectively verify the feasibility, in a clear and understandable manner, of all the steps described in the IFU of the device.

The AI suggests additional questions, including the frequency of potential errors, the presence or absence of a troubleshooting section in the instructions, and an overall satisfaction rating.

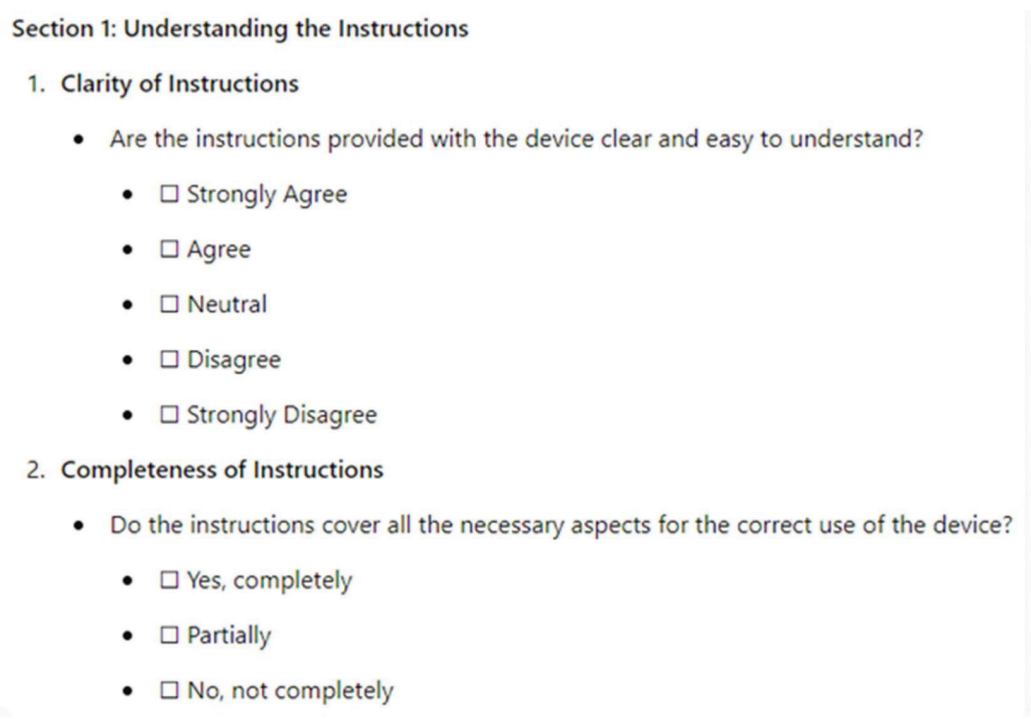
Therefore, in response to generic and unfocused answers, it was necessary to specify the need with a second prompt.

Prompt (2nd)

"Can you write the questionnaire with a more specific approach, that is, include questions that follow the step-by-step use of the device to objectively determine if the IFU are generally clear enough to prevent misuse?"

In response to this prompt, ChatGPT was able to pose questions related to the preparation for use, including the identification of all device and packaging components, and aspects of the actual use of the device, such as the clarity of each step described in the instructions and the ease of execution of the main operation intended with the device (Fig. 8). The general usability evaluation questions and possible improvement suggestions remain the same as those from the first prompt.

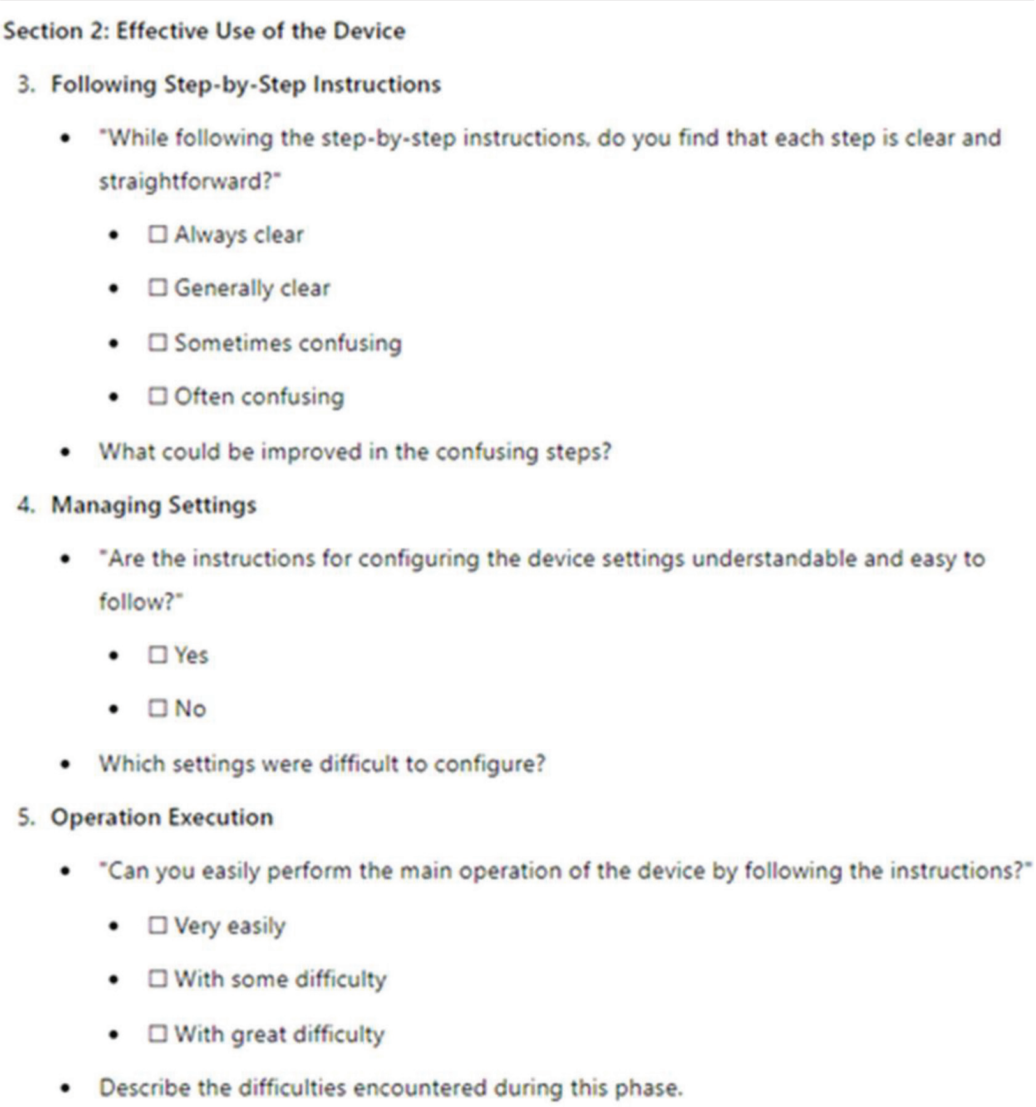
The inclusion of the second prompt resulted in the loss of some responses obtained with the first prompt and the introduction of new ones. Even if the integration of the two outputs were considered, the expected result would not be



Section 1: Understanding the Instructions

1. **Clarity of Instructions**
   - Are the instructions provided with the device clear and easy to understand?
     - ☐ Strongly Agree
     - ☐ Agree
     - ☐ Neutral
     - ☐ Disagree
     - ☐ Strongly Disagree
2. **Completeness of Instructions**
   - Do the instructions cover all the necessary aspects for the correct use of the device?
     - ☐ Yes, completely
     - ☐ Partially
     - ☐ No, not completely

FIGURE 7 - Output of ChatGPT as a question from the usability test after the 1st prompt.



Section 2: Effective Use of the Device

3. **Following Step-by-Step Instructions**
    - "While following the step-by-step instructions, do you find that each step is clear and straightforward?"
        - ☐ Always clear
        - ☐ Generally clear
        - ☐ Sometimes confusing
        - ☐ Often confusing
    - What could be improved in the confusing steps?
4. **Managing Settings**
    - "Are the instructions for configuring the device settings understandable and easy to follow?"
        - ☐ Yes
        - ☐ No
    - Which settings were difficult to configure?
5. **Operation Execution**
    - "Can you easily perform the main operation of the device by following the instructions?"
        - ☐ Very easily
        - ☐ With some difficulty
        - ☐ With great difficulty
    - Describe the difficulties encountered during this phase.

FIGURE 8 - Output of ChatGPT as a question from the usability test after the 2nd prompt.

achieved, namely usability that is compliant and comprehensive according to regulatory requirements.

It was decided to continue querying ChatGPT by providing additional knowledge, specifically an explanation from a regulatory expert on what a usability test for a medical device is and its purpose.

Prompt (3rd)

“Please be much more specific and detailed and delve into the individual functionalities of the device. Before placing the device on the market, the medical device manufacturer must ensure that the IFU provided with the device are clear to the user. To do this, the manufacturer subjects the device to a test by selecting 15 volunteers, who are the actual users of the device. These can be ordinary people or professionals in the case of devices that can only be used by them. Performing a usability test means providing the user with both the device’s IFU or the user manual, as it is provided when the device is marketed, and a questionnaire that must detail each step of using the device, from consulting the packaging, opening the packaging, setting up the device, using the device for its intended function, to storing the device, and performing maintenance if applicable. Therefore, when generating the questions, they must cover all phases, not simply asking the user if the described phase is clear, but whether that step described in the instructions is performed as described and achieves the expected result. For example, in the case of the device in question, it is indicated that the device must be rehydrated before use, and instructions are provided on how to proceed with this rehydration. It is necessary to ask the user if the steps indicated for rehydration are clear, simple, and, most importantly, allow the achievement of the final desired volume of the device. Therefore, the questionnaire questions should not only aim to ask if the IFU are understandable but whether the user’s understanding is objective, meaning that the understanding is not based on subjectivity and user interpretation, but the result obtained by performing the step is objective for all users. Are you able to formulate a questionnaire that follows these

© 2024 The Authors. Published by AboutScience - www.aboutscience.eu

guidelines I have provided? Use the attached file (IFU) as a reference."

Despite being provided with guidelines for designing a usability test, ChatGPT merely added questions to those already generated with the 2nd Prompt, focusing on the ease of opening the packaging and the rehydration phase of the device, with or without achieving the desired final volume of the medical device (Fig. 9).

Therefore, the analysis of what ChatGPT produced, whether from a single prompt, a prompt supplemented with knowledge, or a combination of responses to individual prompts, leads to the conclusion that ChatGPT provides

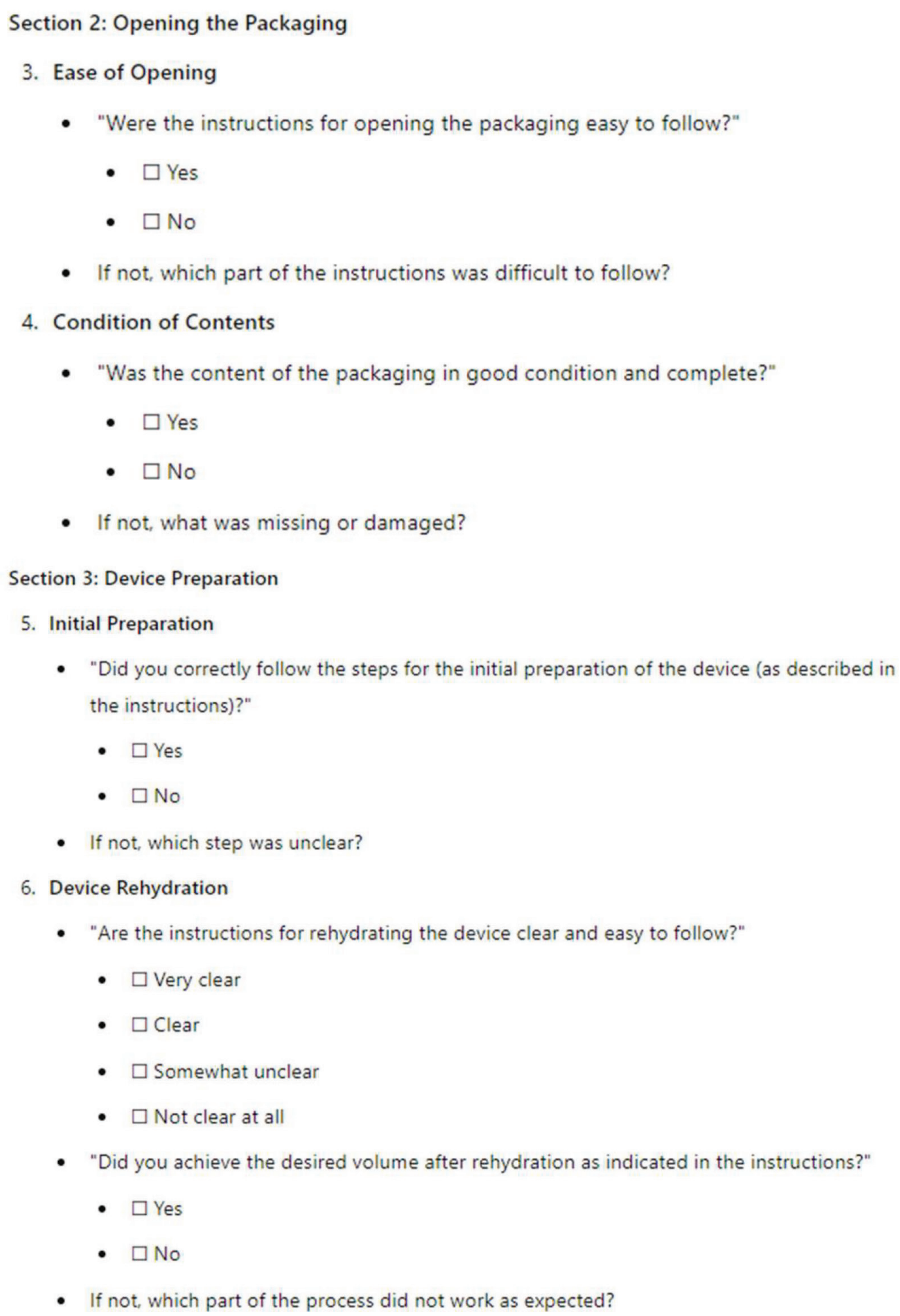


FIGURE 9 - Output of ChatGPT as a question from the usability test after the 3rd prompt.

Section 2: Opening the Packaging

3. **Ease of Opening**
   - "Were the instructions for opening the packaging easy to follow?"
     - ☐ Yes
     - ☐ No
   - If not, which part of the instructions was difficult to follow?
4. **Condition of Contents**
   - "Was the content of the packaging in good condition and complete?"
     - ☐ Yes
     - ☐ No
   - If not, what was missing or damaged?

Section 3: Device Preparation

5. **Initial Preparation**
   - "Did you correctly follow the steps for the initial preparation of the device (as described in the instructions)?"
     - ☐ Yes
     - ☐ No
   - If not, which step was unclear?
6. **Device Rehydration**
   - "Are the instructions for rehydrating the device clear and easy to follow?"
     - ☐ Very clear
     - ☐ Clear
     - ☐ Somewhat unclear
     - ☐ Not clear at all
   - "Did you achieve the desired volume after rehydration as indicated in the instructions?"
     - ☐ Yes
     - ☐ No
   - If not, which part of the process did not work as expected?

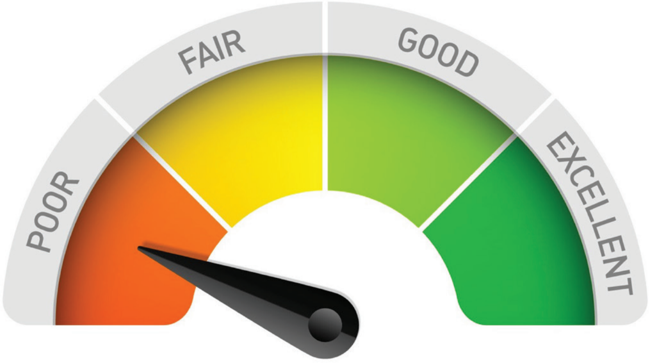


FIGURE 10 - Evaluation Dashboard of ChatGPT's output related to the case study on generating a usability questionnaire for the medical device breast implant.

insufficient support to regulatory affairs in this case (Fig. 10) in terms of relevance, completeness, clarity, accuracy, and comprehensiveness.

Conclusions

The introduction of ChatGPT in the management of regulatory affairs for medical devices represents a significant technological advancement. Thanks to its ability to understand and generate natural language, combined with a sophisticated architecture based on geometric transformations and attention mechanisms, ChatGPT offers a wide range of applications that can complement complex processes, improving operational efficiency and reducing the workload of human operators. This approach translates into support capabilities in regulatory compliance activities, document and report drafting, questionnaire and test creation, data analysis, and scientific information management, increasing productivity and reducing response times.

Challenges and optimization of token consumption

Despite significant advantages, managing token consumption represents a crucial challenge, especially in regulatory contexts where information can be complex and detailed. ChatGPT operates on a model based on token counting, where each word and punctuation mark contribute to the computational cost of generating responses. This can become an issue when long or detailed responses are required, risking reaching the token limits and thus obtaining incomplete responses.

Adopting optimization methods and specific training is therefore essential to fully exploit the potential of ChatGPT. Prompt engineering is fundamental in this context, as it determines the model's effectiveness in understanding and responding to requests. Well-designed prompts not only improve the quality of outputs but also optimize token consumption, reducing the risk of fragmented responses and enhancing the relevance of the provided information.

Continuous improvement through prompt engineering

The practice and refinement of prompt engineering techniques are iterative activities that enable increasingly precise and targeted responses in relevant contexts. Developing detailed prompts, analyzing generated outputs, and training users on how to effectively interact with ChatGPT are key steps to maximize the value of AI in regulatory workflows.

Using context-specific prompts and in-depth knowledge of the model allows requests to be calibrated to extract the system's full potential. For example, for a more technical query, it may be necessary to provide ChatGPT with basic information and context to ensure a comprehensive response, while in other situations, a direct and concise question may be sufficient to achieve the desired result.

Adapting interaction strategies

The case studies clearly show that interaction with ChatGPT must be adapted based on the topic to be developed and the task to be performed. The user is called upon to establish a symbiotic relationship with the AI, gradually understanding the logic of interaction and information exchange. This approach requires continuous commitment to refining interrogation techniques, with the goal of obtaining high-quality outputs and reducing operational times: a simple single prompt may suffice for some generic requests, while other situations may require a sequence of complex prompts, accompanied by detailed technical information, possibly drawn from experts in the field. This approach allows ChatGPT to operate more comprehensively, making the best use of the available information.

Evolution of the model and access to information

The effectiveness of integrating ChatGPT into regulatory processes is closely linked to the continuous evolution of the model, the availability of up-to-date information from the web, and the development of increasingly advanced releases and algorithms. The AI's access policies to informational databases, the expansion of its capabilities through integration with new data sources, and the continuous updating of its knowledge are crucial factors for maintaining high performance levels.

In conclusion, the use of ChatGPT in regulatory affairs for medical devices not only represents a powerful tool to support decision-making processes but also requires a strategic and adaptable approach from users. Continuous training, prompt optimization, and adaptation to technological evolutions are essential elements to maximize the benefits of this technology, enhancing operational effectiveness and ensuring a more efficient and responsive management of regulatory affairs.

Disclosures

Conflict of interest: All authors declare no conflict of interest.

Financial support: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

© 2024 The Authors. Published by AboutScience - www.aboutscience.eu

References

1. Regulation (EU) 2017/745 on medical device. Online. Accessed September 2024.
2. IMDRF MDCE WG/N56 on Clinical Evaluation. Clinical Evaluation | International Medical Device Regulators Forum. Online. Accessed September 2024.
3. IMDRF/GRRP WG/N47. Essential principles of safety and performance of medical devices and IVD medical devices. Online. Accessed September 2024.
4. MEDDEV 2.1/3 Revision 3. Online. Accessed September 2024.
5. MEDDEV 2.7/1 Revision 4**.** Online. Accessed September 2024.
6. MDCG 2020-6. Online. Accessed September 2024.
7. IEC 62366-1:2015 Medical devices – Part 1: Application of usability engineering to medical devices. Online. Accessed September 2024.
8. FDA – Good machine learning practice for medical device development: guiding principles. Online. Accessed September 2024.
9. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. arXiv:1706.03762. Online. Accessed September 2024.
10. Di Bello F Unraveling the Enigma: how can ChatGPT perform so well with language understanding, reasoning, and knowledge processing without having real knowledge or logic? AboutOpen, 2023;10(1):88-96. CrossRef