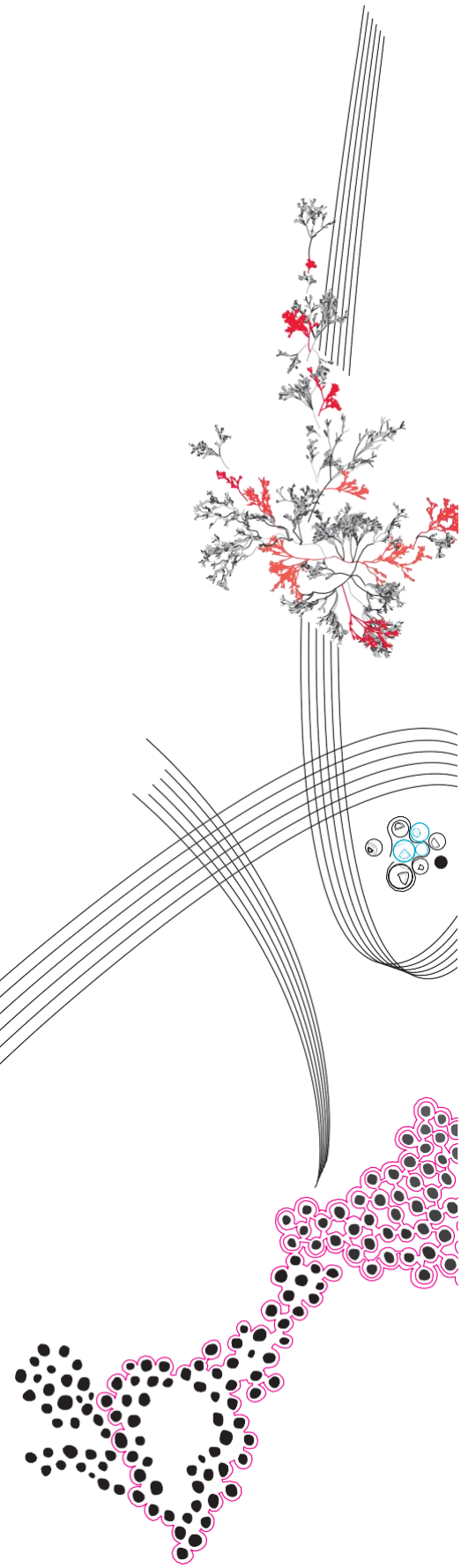


MSc Psychology

Master's Thesis

Modern solutions: Is a Chatbot System for Reporting Medical Device Issues a Promising Alternative to Currently Used Form Systems?

by

Maria Hristova

Supervised and Edited by: Dr. Simone Borsci

Second Supervisor: Dr. Funda Yildrim

November, 2023

Department of Technology, Human and Institutional Behaviour

Faculty of Behavioural, Management and Social Sciences

Psychology

Univeristy of Twente

UNIVERSITY OF TWENTE.

Contents

Abstract

Introduction. This work explores the possibility of utilizing chatbots, powered by natural language processing (NLP), as systems for support during the error-prone task of reporting adverse events (AEs) with medical products. It emphasises the importance of involving of healthcare professionals and patients in reporting AEs to enhance detection rates and address the low levels of incident reports being submitted for review. The present work supports the feasibility of further research into the topic of developing a chatbot-based substitute of the system. *Method.* Using the UK system known as the Yellow Card (YC) system as a base two new systems were derived for testing the assumptions of the study. In collaboration with other researchers from the Imperial College of London a fill-in form system was developed additional to a brand new chatbot-based system. A between-subject experiment to compare the usability and information-gathering capabilities of the two system versions was carried out. *Results.* The findings of this study support the assumptions that by using a chatbot-based system there will be no loss of information in successfully submitted reports and no significant difference in the satisfaction reported by participants when compared to a form. *Discussion.* The paper concludes that using a chatbot system has the potential to address existing issues in pharmacovigilance, namely the low reporting rate and the possible loss of information when incidents are reported, and therefore lead to improvements in the reporting of medical incidents. However further research on the topic is encouraged to confirm the feasibility.

Keywords: Pharmacovigilance, AE reporting, Chatbots, User Experience, Human Factors Engineering, Yellow Card system, Medical Incidents Reporting.

Introduction

The Pursuit of Better Pharmacovigilance

Over the last decade, one of the objectives pursued in pharmacovigilance, which is the science and activities relating to the detection, assessment, understanding and prevention of adverse f adverse effects or any other medicine or vaccine related problem (WHO, 2023), has been to delve deeper into how detecting, assessing, understanding and preventing adverse effects can benefit from patient self-reporting. This idea has been previously neglected, but more and more countries across the globe have either started research into the possibility for both patients and healthcare professionals (HCPs) to report adverse events (AEs) or have already introduced some system that is open to the public Sales et al., 2017 (Hazell & Shakir, 2006; Inch et al., 2012; Sales et al., 2017). An adverse event could be any untoward medical occurrence in a patient or clinical investigative subject who has used a pharmaceutical product. It is not necessary for the event to have a causal relationship with the treatment, as long as the event can be temporally associated with the use of a medicinal product, whether or not considered related to the medicinal product (Gagnon et al., 2012). Involvement from both HCPs and patients can ensure that no such event goes unnoticed, as it allows for both parties to report suspected adverse reactions and increase the overall detection rates. Systems have been deployed in different countries to combat under reporting by covering as many people as possible - patients, HCPs, suppliers and manufacturers (Abrantes & Cordeiro, 2018; Avery et al., 2011; Kiguba et al., 2023; Palaian et al., 2010).

Over the course of the last ten years, the U.S. Food and Drug Administration has received a similar number of reports each year submitted by HCPs and consumers, with five of the years having more reports submitted by consumers (U.S. Food and Drug Administration, 2023). However, this is more of an exception and looking at countries such as the United Kingdom, the Netherlands, or Malaysia, the percentage of reports submitted by non-healthcare professionals is significantly lower (Kiguba et al., 2023; Tase et al., 2021; Van Grootheest et al., 2003). Experts from the pharmaceutical field and researchers have taken up the challenge of exploring the reasons behind those statistics (Lewis & McCallum, 2020), with the final goal of understanding how to further promote to patients the importance of submitting AE reports.

In the UK, in particular, it is estimated that approximately only 1% of all patient safety-related incidents are reported, despite the efforts put into bringing awareness to the importance of reporting medical incidents (MHRA, 2023). Out of those cases, approximately 3% relate to incidents with medical devices. Such a low percentage of reports is severely insufficient for providing encompassing safety of medical devices, which are a market that not only has seen an incredible demand increase over the last 3 years but is expected to grow even more, leading to more uses of medical devices all across the medical field (TBRC Business Research, 2023). The efforts of the Medicines and Healthcare Products Regulatory Agency (MHRA) to higher the adverse events reporting rate have led to research into their system used for AE reporting and there has been work done by Tasse et. al. (2021, 2022) for researching possible setbacks the system is facing and finding out ways for further improvements to combat under-reporting, which is what the current paper will be concerned with.

Main Issues Faced when Reporting AEs

Looking at the system provided by the MHRA through the human factors lens has given a novel approach to identifying issues that are hindering both healthcare professionals and regular users (Tase et al., 2022). Hopefully, the solutions to those concerns, considering that other industries have successfully adapted HFE principles into their practices. With digitalization becoming a big part of the economy, chatbots have become an integral part of supporting the influx of digital interactions (Li et al., 2023). This has also reached the reporting of adverse events and the UK government now supports a website where everybody can submit their report about an adverse event they or somebody they know has experienced using the Yellow Card (YC) system.

As of the beginning of 2023 on the website of the YC system provided by the MHRA, 26 different case studies have been published, all of which were initiated because of submitted reports, but only a few relate to medical devices (MHRA, 2023). This is direct evidence of the importance of AEs reporting as it shows how the MHRA utilises the information given to improve safety and work on the detection of unsuspected hazards. One of the cases is the case of Continuous Glucose Monitoring Systems causing skin irritation that has led to the manufacturer working closely with MHRA to ensure the issue was captured and assessed adequately. With the complex concept of safety in medicine, there is a need for a detailed reconstruction of the environment that led to the incident when a report is submitted in case there are unforeseen factors (Carayon et al., 2014; Liu et al., 2021). In the current YC system, one of the leading factors that have been pointed out as insufficient is the specificity of the

recorded reports (Tase et al., 2021). A different and targeted approach to malfunction reporting of MDs is required with an emphasis on effective communication to ensure better performance and motioning of patient safety.

The form needs to be able to adjust for medicines, medical devices, vaccines and other health- related cases. The YC does have a different set of questions that need to be filled in for overarching categories, but it does not differentiate on individual level between different subjects (MHRA, 2023). For example, the questions asked for a one-time use deice would be the same as for a device that is used daily, while it might be of value to acknowledge the difference. Moreover, the YC system has been reported to have a complex structure that does not encourage use, even by medical professionals, and the issue is exacerbated when a non-medically trained person encounters it (Tasse et al., 2023). The system itself does not provide any support to the user on how to obtain certain information, for example, the name or date of expiration of the device, nor prompts specificity when describing the incident (MHRA, 2023). Additionally, complexity of the system might be unintentionally preventing from providing detailed information about the event being reported and therefore decreasing the quality of the reports. Addressing those aspects of the reporting system and providing a solution that could be used to increase both the amount of reported information and the ease of doing so will ultimately lead to a better knowledge of incident causes and will impact patients' safety long term.

History of Chatbots in Healthcare

Appropriating the human-centered approach and co-design strategies relating to the detection, evaluation, understanding, and prevention of adverse reactions to medicines or any other medicine-related problems by including the patients, as well as the HCPs, has great potential in improving the perception of the importance of pharmacovigilance in the general public's eye (Turk et al., 2022). Instances of introducing advanced technological solutions to improve aspects of the system can be seen all over the healthcare industry. From utilizing a chatbot-based mobile application for human mental health (Podrazhansky et al., 2020), to harnessing the power of data gathered from social media to improve pharmacovigilance (Pappa & Stergioulas, 2019).

Looking towards other industries such as e-commerce or the automotive industry, one of the most widespread technologies used to implement virtual assistance nowadays is chatbots (Reshmi & Balakrishnan, 2018). Chatbots are systems that have the capability of responding to a variety of inquiries tailored to the specific use case through the use of Natural Language

Processing (NLP) (McNeal & Newyear, 2013). Modern chatbots have been used as virtual assistants for training, systems for customer support, or simply put up to perform routine tasks efficiently (Casillo et. al., 2020; n.d, 2023). If undisclosed whether there is a human respondent or a chatbot on the other side of the query, there are even some systems that can pose as humans (Hazell & Shakir, 2006). Conversational agents can be contextually aware and access vast pools of data with immense speed. Closed-domain tasks, such as banking, have seen a successful attempt to train a contextually aware chatbot to direct the user quickly and effectively to the right resources (Bhattacharyya et al., 2020). The success of implementing some level of automated assistance, in those and many other domains, can speak for the feasibility of introducing such technology to AE reporting, especially because there are very strict regulations in place about medical products, giving a perfect opportunity for a system to be trained on all necessary details.

A systematic literature review on the topic of chatbots in healthcare has revealed conversational agent-based interventions are a feasible and acceptable solution, and have had positive effects on physical functioning, healthy lifestyle, mental health and psycho-social outcomes (Li et al., 2023). Considering chatbots have also proven to be a great tool for information extraction (Martinez et al., 2020), it is not unthinkable to believe they might be helpful when addressing issues with AEs reporting. With the help of natural language processing (NLP), chatbots are able to turn the natural language of a human into structured information and make use of it (Chowdhury et al., 2003). Given there is an appropriate data pool, a chatbot can be trained to be very well knowledgeable in certain fields, for example determining whether an individual has a mental illness based on text interactions and suggesting prevention methods by using machine learning algorithms (Podrazhansky et al., 2020). Specifically for data extraction, a chatbot has been proposed to be used as a solution for data extraction from users' queries and has proven to function as a reliable and transparent tool for effective communication for both patients and medical staff (Mittal et al., 2021). Moreover, chatbots are a great tool for lowering the accessibility threshold, since they can be voice activated and provide assistance to the user if needed (Li et al., 2023).

Aims of the Present Work

With the help of conversational AI it would be possible to reform the YC system in such a way that it could have specific knowledge about each individual device, medicine or vaccine and increase the amount of specificity and therefore usefulness of the information in each submitted report (Vasquez et al., 2018). Next to that, conversational AI could be trained

to "help" users when they are reporting an AE but have questions, for example, where a specific piece of information about their medical device can be found, which will ease the process of submitting the report, making it more desirable to revisit the system. With the vast capabilities of chatbots, it could even be that given the proper tools, people could report information verbally, making the entire YC system more accessible (Feine et al., 2019). However, training and introducing such a tool is both an expensive and extensive process. Ensuring that there would be beneficial outcomes is not completely possible, but what can be done is to explore whether a chatbot system can be a viable alternative to the current form by performing a simplified experiment. To address this, the perception of the users on how usable a chatbot system is for submitting a report needs to be considered. The System Usability Scale (SUS) is a quick and easy scale designed for measuring usability (Hyzy et al., 2022). This scale is short, easy to administer and has been validated for scientific use for estimating the usability of website tools, but it has been also successfully applied to chatbots and it has shown comparable performance with scales designed to test specifically chatbot systems' validity (Borsci et al., 2021).

In the present study, we intend to explore if a chatbot system for reporting medical device issues is a promising alternative to current online form systems; therefore, we will compare people's experience with the YC system (online form) and a chatbot-based system for reporting incidents related to, for instance, medical devices, medicines, vaccines and e-cigarettes. It should be highlighted that we will compare a prototype chatbot with minimal functionalities versus a replica of an online form system that is fully functional and based on several years of validations. We intend the chatbot prototype as a manifestation of an idea, in line with Lim et al. (2008), specifically we are testing for generating and exploring in a low-cost way future steps of development.

Under this perspective, reasonably we do not expect a superiority of the (prototype of) chatbot system versus the replicated current system. Nevertheless, if by using the two modalities of reporting, people will perform and be satisfied without significant differences, this would suggest that a low-level functioning chatbot system is comparable in terms of experience to the current online form system. If such system, albeit possessing only simple functionalities is comparable in those qualities, it is a suggestion for future investment in developing the prototype even further or searching for other alternative approaches. Ultimately, a chatbot could be a powerful tool when applied in the right way. With the help of modern NLP techniques chatbots can be useful in information extraction and with AI advances it is possible to even train systems to probe for specific information and that way be more

useful in combating both under-reporting and lack of content in reports. Based on this the following expectations are posed as central research questions to this work:

1. People reporting AEs by using a chatbot or an online form will have comparable levels of satisfaction measured by the SUS scale, independently from the level of difficulty of the scenario.

2. People reporting AEs by using a chatbot or an online form will have comparable performance in terms of the correctness of the submitted report i.e., effectiveness.

To achieve the purpose of the present work, it is necessary to develop a functional version of a chatbot as well as a replica of form-based version of the YC system. Therefore, this work will have two phases. The first phase will entail the development of a chatbot and fill-in-form-based YC systems. Afterwards, in the second phase, an experimental study will be carried out to compare the two reporting systems.

Phase One: Designing a Chatbot-based Reporting System for Testing

To effectively test the hypothesis this study poses, there needs to be an appropriate study design that will allow for it. Establishing the differences, or lack thereof, in the usability of a chatbot and a form for reporting medical incidents would need a direct comparison between the two to be made. As of the current point in time, there is no officially released chatbot equivalent of the YC system and therefore another

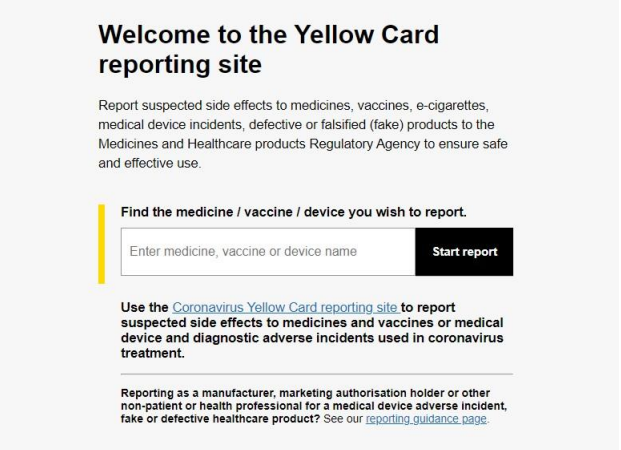
Building a functional chatbot

As a potential replacement for the form format of the YC system, ultimately a chatbot needs to be able to detect and understand the intentions of the user appropriately and provide the support required to obtain a comprehensive and specific report. Previously, research has been carried out on the YC system using a fill-in form that mimics the real system flow to collect data (Tase et al., 2021; Tase et al., 2022). The research team, affiliated with the Imperial College in London have kindly provided their materials to be used as a base for developing a test chatbot system. Taking a look at the official landing page of the YC system that the public can access (MHRA, 2023) reveals that when the system is presented for the first time, users are given the option to either start the report by typing a product name or be directed towards a manual that would explain to them how to fill in the report based on their role in the supply

chain for medical products as shown in Figure 1. From this point forward the reporting system has several different flows that can be triggered, depending on the type of medical product and the position of the person reporting the incident (patient or HCP).

Figure 1

First encounter with the YC reporting system for users.



Proposed chatbot design

Since the form provided by the research team of Tasse et. al. (2022) is constructed to collect a report for medicines and focuses on the healthcare provider side, some adjustments were made to it first. In the YC system, several questions are shown only when reporting adverse events with medical devices such as requiring serial numbers of the device. Those questions were added to the base form and the modified structure was used as a model for the test chatbot system. Some of the other data required includes for example the age and gender of the person involved, details about the incident and the current status of the person. The whole report can be divided into four separate sections: information related to the reported product, information about the experienced side effects, information about the person involved and information about the person reporting the incident. In each category, the information that the form considers essential and does not allow for further progression with the report unless provided with the information labelled to be mandatory as it can be seen in Table 1.

Table 1

Information Required for ADR Report According to the MHRA

Type of information	
Mandatory	Optional
Product information	
Type of product Product name Current condition of the device	Manufacturer / Supplier Device serial number
Incident information	
Description of the incident in own words	Severity of incident
Patient information	
*	Age, gender or name
Reporter information	
Affiliation with healthcare Status as a reporting person	Contact information for further investigation

Note. This table contains the different types of information requested to file a report using the YC system when reporting about a medical product malfunction, both mandatory information and additional information that can be provided for a more complete report.

** Giving at least one of the information units (Name, Age, Gender) is sufficient for the YC system to accept the report.*

However, one major restriction that needs to be acknowledged is that the chatbot system developed for this study does not allow for information to not be filled in, regardless of whether it is optional or not in the form equivalent. On one side this can greatly contribute to obtaining a more detailed report about the ADR, but at the same time, this limits the way interaction with the system can be carried out. In this study, it is a byproduct of the way the platform chosen for building the conversational agent works and not a conscious choice for the design. In similar way, entering unrelated to the scenario data is not properly handled by the system – there is no way to make distinction between some types of information, such as different date, names, and address. On the other hand however, the system very is very rigid with other types of information – for example the product being reported (medical device), or the name of the medical device, and will not function properly given random input there. With more advanced prototype this can be easily overcome, but for all intents and purposes of this

study the system behaves in an appropriate manner and participants are strongly encouraged to not stray from their given scenario.

The system was designed using the DialogFlow ES platform for natural language processing (nd, 2022). Following the general outline of the form and transforming the questions into more interactive ones that can be posed by the chatbot a conversational flow was designed. To ensure that users are given the opportunity to fill in the same information as in the form version, requests for each unit of information as seen in Table 1 were integrated into the flow of the system. Each unit of information was included as a separate entity that the chatbot recognises and tries to obtain, more than 100 test conversations were run to ensure proper training and recognition of contextual cues and phrases the user could possibly use.

Validation of Design

After the conversational agent was created and trained by the research team, a small pilot test was conducted. Pilot testing is an important phase for it can ensure early detection of possible issues with the system that were missed by the researchers, as well as provide a novel view over what could be improved before data collection to enable easier interaction (van Teijlingen & Hundley, 2002). Five volunteers were recruited to participate via personal networking. The testers were instructed to fill in a report and then a short free-form interview was conducted with each one to talk about their experience with the system.

Pilot Testing Outcomes

One out of the five people participated fully remotely, and the rest were asked to perform the task in the presence of the researcher, in case the system malfunctions. Three people managed to successfully complete their report, one person almost managed but the system malfunctioned on the last question, and one person was halfway through the report when the system malfunctioned and did not want to proceed further. Overall, all of them reported liking the interaction with the system. A point that was brought up in the follow-up conversations more than once was the rigidity of the system and how certain phasing seemed to be unrecognisable by it despite being correctly associated in the context. For example, the system refused to accept " glucose monitor", but accepted " blood sugar monitor" as a valid answer to "What product are you reporting about?" despite the two phrases having the same meaning. This extended to the chatbot struggling to recognise properly when names, age or gender were given as input, particularly when the user wanted to input all three of them in the

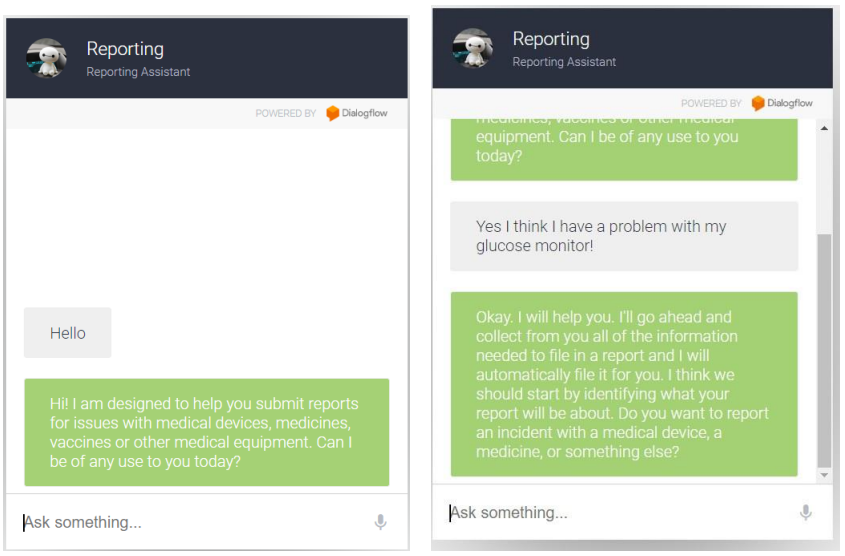
same message. Another point that was brought up was that it would be nice if the system could provide guidance on where to find certain information, for example, LOT number or SN number, as well as help on how to find the right date if you are looking for date of manufacturing or expiry date.

Final Version of the System

After the pilot test of the system new phrases were introduced to the chatbot and additional training was carried on. This led to an improvement in the overall quality of interaction. On average, a full report using the system can be submitted with a conversation of around 30 back-and-forth messages between the user and the chatbot. Because of system limitations, the agent is not capable of initiating a conversation on its own and needs to be prompted by an opening message such as "Hello", "Hey", etc. At the beginning of the conversation with the system, as it can be seen in Figure 2, the system explains its purpose and then the user is given options on how they want to proceed, similarly to the YC form system.

Figure 2

Initiating a report with the chatbot

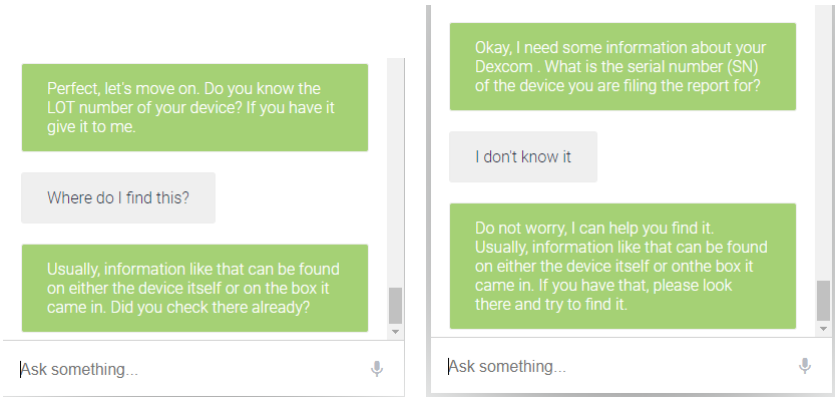


Note. This figure depicts how interaction is initiated between the user (in grey) and the chatbot (in green) from left to right.

Taking into account the feedback from the pilot testers, a function that helps users locate information upon request for help was introduced to the system. Separate segments for helping with finding the name of the medical device, the SN and LOT number, and the differentiating between the dates of manufacturing and expiry. Exemplary phrases that could trigger the help prompts can be seen in Figure 3.

Figure 3

Phrases that trigger help prompts from the chatbot



Note. This figure depicts examples of requests for help with identifying information between from user (in grey) and the chatbot's replies to them (in green).

Phase Two: Comparative Study For Testing the Chatbot-based Reporting System.

Methods

Materials

Scenario

To test the capabilities of the two systems, participants were asked to file a report acting as a person who has experienced a malfunction of a medical device. A scenario describing the incident and providing some additional details about the person was specifically created for this study.

The scenario used describes an incident with a blood sugar monitoring system (see Appendix A). The incident was purposefully designed to have no severe consequences for the imaginary person to avoid possible negative connotations for the participants. The imaginary monitoring device used in the scenario is modelled after a real system to ensure all functionalities are realistic, but the information was changed.

Reporting systems

Two systems for reporting issues with a medical device were developed. One of the systems is chatbot-based and the other one is form based as follows:

- The Chatbot-based Reporting Systems (CRS): The chatbot that was designed for this experiment was created with the help of Dialogflow (see Appendix B). Due to technical limitations, the chatbot was created in a way that allows it to be functional only if presented with information related to the specific device in the scenario. It mimics the YC flow that a user would encounter if they were to use the system to report the same incident, however, it has added functionalities such as being able to provide help if needed. Participants were presented with a link to a website where they could interact with the chatbot.
- Form-based Reporting System (FRS): A form that was modelled after the YC system was used to collect data for the control condition. In earlier research on the YC system and how it is perceived by experts who use it, a similar form was created to collect data about the perceived usability of the YC system (Tase et al., 2021; Tase et al., 2022). This form was kindly provided by the original researchers and adapted to the

purpose of the current study by slightly altering the flow to be in line with the scenario (see Appendix C). The new version was used in the control condition to represent the YC system in the survey environment.

Procedure

Prior to launching, this study was approved by the ethical committee of the University of Twente with approval number 230414. To collect data from participants, a questionnaire was created using the Qualtrics platform. Upon first encounter, the participants were informed of the purpose of this study and what they could expect, after which they were asked for their informed consent. In case participants did not consent for their data to be collected and used for research purposes, they were directed straight to the end page of the survey and their responses were not recorded in the database.

Consenting participants were asked some demographic-related questions, such as age, proficiency in English, occupation, etc. Next, each one was randomly assigned to one of the two reporting methods: chatbot or form, in combination with randomly receiving the low or high difficulty time constraint version. After completing the task of submitting a report given a prepared scenario for this study, participants were asked to fill in a System Usability Scale and rate the 10 statements from "completely disagree" to "completely agree". Participants who interacted with the chatbot had an additional question to control whether they succeeded in completing a report or not and if they didn’t, some clarifying questions about their interaction were included as well to probe for possible pitfalls of the system.

The design of the study suggests a length of between 10 and 20 minutes for participants, depending on the condition they are assigned to. The participants can access the study, which is hosted on Qualtrics, through unique links and there they will find the questionnaire and all additional materials they will need to complete the study. Because healthcare and medical device and medicine safety is an important topic and should any of the people who were involved in the experiment experience issues in real life at any point after, it is crucial to ensure they are aware of the nature of this study and that it does not substitute the systems in place provided by the government. Throughout the study at several points and in the end, participants were reminded that the systems they encountered are not to be taken as official reporting systems and do not serve any other function apart from supporting this research. Participants were also reminded should they experience any issues related to pharmacological safety; they should seek advice from their official healthcare provider.

Design

The present study is designed as a between-subject experiment with participants randomly assigned to file a report about the same AE with a medical device with either a chatbot-based reporting system (CRS) or form-based reporting system (FRS). For each condition, participants were also randomly assigned to be presented with one of two time constraints for submitting the report to regulate difficulty - low difficulty (16 minutes) or high difficulty (8 minutes). Since not all reports are submitted immediately after the reported AE, there is a possibility for memory decline upon recollection. This can influence the recall of detailed information, and introducing time constraints in the study design is done to a lot for different times for recall, simulating the real passage of time or sense of urgency when reporting. This way the information can be more representative of all types of reporting situations - immediately after the AE or some time afterwards.

To summarise, this is a 2 (CRS/FRS) * 2 (low/ high difficulty) factorial design. Two aspects are going to be used to compare the two reporting systems under different levels of induced difficulty: i) Report correctness and, ii) reported perceived system usability, as calculated by the overall SUS score.

Participants

A total of 32 participants (16 Female, 15 Male, 1 Non-binary) with a mean age of 24.62 (SD: 7.61) between the ages of 18 and 53, consented to partake in the study. Twenty were randomly assigned to the FRS conditions and 12 to the conditions that used a chatbot, with 16 people in total assigned to perform under high-difficulty time constraints.

The experiment was performed in English. Twenty-one participants indicated they have advanced proficiency in the English language, four people reported English as their native/primary language and seven people said they have an "intermediate" understanding of the language. Predominantly the sample consisted of students (21 participants), and only 4 participants indicated that they were related to the healthcare industry, either by educational background or current position.

Data Analysis

The survey was distributed through Qualtrics XM and the same system was used to record the participants' answers. The data was later exported and prepared for analysis in

Excel, after which it was imported into R Studio. Data from participants who did not give their consent was not recorded by the system at all. Participants who had not completed the survey in full, meaning they did not provide answers to all of the 10 SUS items, or in case they were assigned a chatbot condition and did not interact with the system at all were removed. For each participant, the SUS score was calculated by first standardising the scale and accounting for positive and negative answers, after which the scores from each item were summarised and multiplied by 2.5.

Additionally, for each participant, the correctness rate of the mock report they submitted was calculated. The end result was the percentage of correctness for each report, given they have fulfilled the requirements to provide all of the essential information for a report to be considered valid. In the FRS condition, if a participant had not provided information to one or more of the five fields required for a valid report, their score was immediately set to 0% and their report was considered incomplete, resulting in exclusion from further analysis. For the participants who used CRS, if a participant has placed wrong information in any of the mandatory required fields, their entry was also considered invalid for use. The maximum obtainable score for fully completing the report and providing all of the key details was 20 points for the form condition and 19 points in the CRS condition (see Appendix D and Appendix E).

As one of the questions required participants to provide information about the incident in their own words, this question was separately evaluated qualitatively to establish a point value from 0 (not providing any relevant information from the AE scenario) to 5 (providing detailed information about the AE and surrounding details) with one point given for each of the following:

- provides information related to the given scenario;
- relays that there is a loss of connection between one part of the CGM system and the mobile application
- specifies the frequency of connection loss;
- provides information about other factors such as phone state or specific information about
- the setting in which the accident occurred;
- provides information about the aftermath (was the connection ever restored, what was attempted to restore the connection);

For both conditions, the entrees were evaluated on whether they gave information to any of

the five aspects listed, with giving relevant information to the scenario being mandatory. If a participant has given a description of the situation that does not relate to the original scenario provided, their answer was given 0 points but still was considered a valid answer. The specific scoring criteria with all of the data entries that were used can be found in Appendix F. Other additional materials can be found in Appendix G.

Results

Data Preparation

After completion of data collection, the data was exported from Qualtrics into R and was prepared for data analysis. Participants who did not provide complete responses to the SUS items were removed and their data was discarded. After this, the total number of participants across the four conditions was 39, 20 of which used a form to submit a mock-up report and 19 used a chatbot. Upon further investigation of the participants in the chatbot condition, the logs obtained from four of the participants did not contain any relevant to the survey information and after deliberation, their survey entries were disposed of. The final number of participants used for analysis was 32, with 12 participants in the low difficulty control condition, 8 in the high variation and 4 and 8 participants respectively for the chatbot conditions.

In each condition, regardless of the level of difficulty, a variable "correctness" was created to represent the percentage of data included in the report by the participant. The percentage was calculated based on whether the participant reported any information at all, if it was in line with the scenario and to what level were details included in the description of the incident. The total possible score for the FRS was 20 points, and 19 points for the CRS (see Appendix D).

Manipulation check

Table 2 summarises the number of participants, the average SUS score and the correctness of response using the different systems at different levels of difficulty.

Table 2

Summary of participants, average SUS and correctness scores per condition

Type of Reporting System		Low Difficulty	High Difficulty
Form (FRS)	Number of Participants (n)	12	8
	Average SUS Score	75.62	73.44
	Average Correctnes	74.17	79.44
Chatbot (CRS)	Number of Participants (n)	4	8
	Average SUS Score	70.62	82.50
	Average Correctnes	66.50	85.32

To check if our manipulation of the time constraints (difficulty) affected the dependent variables we performed multiple pairwise t-test analyses between various combinations of assigned reporting methods and difficulty. The results indicated no statistically significant differences among the group and difficulty level combinations for 'correctness' (all p-values > 0.05). A similar analysis was conducted for the 'SUS score' variable, revealing no significant differences across the group and difficulty level combinations (all p-values > 0.05). The use of the Benjamini-Hochberg method helped control for multiple comparisons in this analysis.

Effect of reporting method and difficulty on users' satisfaction

A two-way robust ANOVA was performed to evaluate the effects on user satisfaction (SUS) due to the usage of one of the two reporting systems (From/Chatbot) and the two levels of difficulty (Low or High) in line with ***RQ 1***. The means and standard deviations for SUS scores per group are presented in Table 3 below.

Table 3

Descriptive statistics for SUS scores

Reporting System	Difficulty	*M*	*SD*
Form (FRS)	Low	75.63	19.57
	High	73.63	20.99
Chatbot *(CRS)*	Low	70.63	9.87
	High	82.50	*14.40*

The results from a two-way ANOVA with robust estimators indicated no significant main effect for reporting type, F(1, 28) = 0.334, p = .568, partial η^2 = .01; no significant main effect for difficulty, F(1, 28) = 0.186, p = .670, partial η^2 = .01; and no significant interaction between reporting type and difficulty, F(1, 28) = 1.050, p = 0.314, partial η^2 = .04.

Effect of reporting method and difficulty on correctness of reports

To test **RQ 2**, a robust two-way ANOVA was performed to evaluate the effects of the correctness of participants (scores in percentage) in reporting using one of the two systems, and considering the level of difficulty**.** The means and standard deviations for the correctness score are presented in Table 4 below.

Table 4

Descriptive statistics for report correctness scores

Reporting System	Difficulty	*M*	*SD*
Form	*Low*	74.17	30.46
	High	79.63	17.70
Chatbot	*Low*	66.50	44.56
	High	85.37	6.84

The results from a two-way ANOVA with robust estimators indicated no significant main effect for the reporting method, F(1, 28) = 0.084, p = .774, partial η^2 = <.01; no significant main effect for difficulty, F(1, 28) = 1.176, p = .287, partial η^2 = <.01; and no significant interaction between reporting type and difficulty, F(1,28) = 0.462, p = 0.502, partial η^2 = <.01.

Discussion

The present work aimed to explore the idea of substituting the current YC system for reporting medical with an adaptive assistant and what benefits this could provide to both the users and the responsible authority. This work focused on comparing two possible versions of the reporting system on how well they perform with extracting information from the person reporting and their perception of the usability of each method.

Two systems we tested – CRF and FRS – by being given to participants as a means to submit a mock

report about an incident with a medical device under the manipulation of two different time constraints. Each participant was presented with only one of the available systems and afterwards, they were asked to fill in the System Usability scale to assess their experience.

What resulted from this study was in line with the expectations set in the beginning. Namely:

1. Regarding the satisfaction of the participants, the results suggest that there is no difference when the report is submitted using CRS or FRS, or due to the introduced time constraints.
2. Correctness of reporting is not affected by the usage of CRS or FRS, nor by the level of difficulty we manipulated.

These results, albeit seeming not providing any evidence to support that the ADR process will benefit from a chatbot system, are quite valuable. As this is a study set to evaluate the feasibility of a novel idea in a low-cost setting, the systems performing equally in the eyes of the users is a very satisfactory outcome. When reminded of the idea that chatbot agents are great tools for information extraction (Vasquez et al., 2018), the advantages of such systems become apparent. Despite the CRS used for testing in this research being a very rudimentary one, and despite the fact there are currently way more advanced models that can be employed, in the present study participants using it did not perform worse compared to those that used a traditional reporting system (FRS), nor did they report having experience that was significantly different in terms of their satisfaction.

Overall, these findings suggest that it is not without worth to pursue further developing the current prototype to a more advanced chatbot system and assessing if the usage of a more refined CRS might bring better outcomes, both performance and experience wise, compared to the FRS. Certainly, a more advanced chatbot system might also bring additional features and advantages compared to online form system. For once, recent advances in Large Language models (LLMs), such as GPT, there are endless possible functionalities a conversations system can have. Investing time and effort into further evaluating the use case of chatbot as a system for medical reporting can open the doors to functionalities such as: i) users being able to ask for help with finding information if needed, ii) the system can potentially be voice-activated, and iii) a system can be trained to have explicit knowledge about specific medical products or devices to adapt to the case and help with specific information extraction (Abrantes & Cordeiro, 2018). These advantages might bring people to perform better or to have a better experience with future CRS systems compared to FRS systems, as well as introduce the aspect of inclusivity by giving various options of how one can submit a report.

There have already been successful cases of implementing chatbots and AI solutions in healthcare, such as mental health diagnostics or hospital feedback (Classen et al., 2015; Podrazhansky

et al., 2020) and such previous work could be used to extrapolate lessons learned. The progress conversational AIs have made over the past couple of years is evident in their vast application, such as being used for assistance in cognitive behavioural therapy additions (M., 2023) or mimicking a companion that can help with assistance in curating plans, generating ideas, planning trips, and seeking relationship advice. Contextual awareness is one of the strengths such a system can provide and in the case of AE reporting, it is of utmost importance to have context about the specific medicine or device being reported and the means to procure all possible relevant details to further help with improving pharmaceutical safety in the future by bringing more awareness to possible adverse reactions and being able to provide crucial details (Abrantes & Cordeiro, 2018; Cheng et al., 2019). If made with the intention to do so, such a system will have the capabilities of not only requesting the right type of information for the specific AE case but also will be able to give prompts and assistance based on a contextual quest to help the reporter with finding the right information.

Circling back to the expectations of this study, it seems there is no notable effect of what type of form was used to submit a report on the experience of the user, nor on the information they provided. Exploring more the idea to address current issues found with the YC system used for reporting incidents related to medical devices, medicines, vaccines and e-cigarettes by introducing a chatbot based system holds value for more than one reasons. Exchanging the current form-based reporting system of the YC for an AI-driven one that can be provided to users via a chatbot does not have any immediately detectable drawbacks since there seems to be no loss of information collected in reports as well as no apparent difference in satisfaction scores, regardless of the time given to participants to submit a report.

Limitations and future suggestions

This study has identified several limitations that should be acknowledged, as they may have influenced the results and subsequent interpretations. Further research is needed to address these limitations and enhance the validity and generalizability of the findings.

First and foremost, there is something to be said about the sample size of this study. The end pool of recorded and usable data ended up being rather small given the requirements of the statistical methods used to explore the expectations set in the beginning, and once the data is divided in the respective conditions there is a rather large difference between the number of participants in each, making some analysis impossible because of skewness and introducing uncertainty to the results. Despite initial attempts to include at least 10 participants per condition, the final sample size varied greatly across conditions due to the removal of invalid data entries, with one group having 12

participants and another having 4. This discrepancy in participant numbers may have introduced bias and affected the statistical power of the study. Future research should aim to recruit a larger and more balanced sample to increase the representativeness and robustness of the findings.

A second limitation is associated with the chatbot system used for testing, which was not developed by specialists in conversational agents. Taking into account that the chatbot system used in this experiment, while developed with the help of a lot of literature on the topic of chatbot construction and several pilot testers, is still a system developed for preliminary testing and can be improved upon by professionals in the field, it is reasonable to believe that if this experiment was to be recreated with a better chatbot system, it could produce different findings. Consequently, the system occasionally failed to meet users' expectations, leading to data loss in the study. To mitigate this limitation, future research should involve collaboration with experts in the field of conversational agents to ensure the development of a more reliable and user-friendly chatbot system. This improvement would minimize data loss and enhance the overall quality of data collected.

The third limitation stems from the combined effect of the small sample size and the sample characteristics. The study predominantly included participants of younger age with at least a high school level of education. Moreover, it did not account for the experiences of individuals with various disabilities, such as physical, psychiatric, or mental disabilities. To address this limitation, future research should strive for a more diverse participant pool, encompassing individuals from different age groups, educational backgrounds, and disabilities. This broader representation would provide a more comprehensive understanding of the phenomenon under investigation and ensure the generalizability of the findings across different populations.

While this study has provided valuable insights, it is important to acknowledge the limitations that have been identified. By addressing these limitations in future research, such as increasing the sample size, improving the chatbot system, and expanding the participant pool, researchers can strengthen the validity and applicability of these findings.

Conclusion

Based on this study, a suggestion for further development and exploration of the idea to substitute the current form reporting system of the YC for an AI-driven one that can be provided to users via a chatbot is an intriguing and feasible endeavour. This study shows that there are no immediately detectable drawbacks of doing so and therefore there can be more merit into encouraging development of the topic. The expectations of this study were both confirmed and there seems to be no influence on the difference in the satisfaction of participants or their correctness of reporting an adverse event by the type of system they used, nor by the manipulated difficulty. With the help of the human factors approach to problem-solving, the findings from this

research can prove to be important for evaluating the possibility of new technology for reporting that has the potential to aid users and provide interactive adequate support when and where needed.

Acknowledgements

The authors and editors would like to acknowledge Dr. Arkeliana Tase and Dr. Melody Zhifang Ni for their valuable contributions to this research. Their work served as a foundation for this paper and greatly enriched the findings. The authors would also like to acknowledge the researchers for granting access to essential materials that were used in the study, enabling a thorough investigation. Their support and collaboration were instrumental in the successful completion of this research project.

Bibliography

Abrantes, D., & Cordeiro, J. (2018). Extracting Adverse Drug Effects from User Experiences: A Baseline. *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, *2018- June*, 405–410. https://doi.org/10.1109/CBMS.2018.00077

Avery, A. J., Anderson, C., Bond, C. M., Fortnum, H., Gifford, A., Hannaford, P. C., Hazell, L., Krska, J., Lee, A. J., McLernon, D. J., Murphy, E., Shakir, S., & Watson, M. C. (2011). Evaluation of patient reporting of adverse drug reactions to the UK 'Yellow card scheme': Literature review, descriptive and qualitative analyses, and questionnaire surveys. *Health Technology Assessment*, *15*(20), 1–234. https://doi.org/10.3310/hta15200

Borsci, S., Malizia, A., Schmettow, M., van der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2021). The chatbot usability scale: The design and pilot of a usability scale for interaction with AI-based conversational agents. Personal and Ubiquitous Computing, 26(1), 95–119. https://doi.org/10.1007/s00779-021-01582-9

Carayon, P., Wetterneck, T. B., Rivera-Rodriguez, A. J., Hundt, A. S., Hoonakker, P., Holden, R., & Gurses, A. P. (2014). Human Factors Systems Approach to healthcare quality and patient safety. Applied Ergonomics, 45(1), 14–25. https://doi.org/10.1016/j.apergo.2013.04.023

Cheng, M., Moher, B., Napke, E., Lehtiniemi, L., Erskine, J., & Gaamangwe, T. (2019). *Medical device regulations and patient safety* (Second Edi). Elsevier Inc. https://doi.org/10.1016/ B978-0-12-813467-2.00054-7

Classen, D. C., Pestotnik, S. L., Scott Evans, R., & Burke, J. P. (2015). Computerized Surveillance of Adverse Drug Events in Hospital Patients The most common drug classes involved were analgesics, anti-infectives, and cardiovascular agents. Conclusion.=m-that screening for ADEs with a computerized hospital inform. http://jama.jamanetwork.com/

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Con- versational Agents. *International Journal of Human Computer Studies*, *132*(September 2018), 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

Gagnon, S., Schueler, P., & Fan, J. D. (2012). *Pharmacovigilance and Risk Management*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-415787-3.00013-8

Google. (n.d.). Dialogflow | Google Cloud. Google. https://cloud.google.com/dialogflow

Hazell, L., & Shakir, S. A. (2006). Under-reporting of adverse drug reactions: A systematic re- view. *Drug Safety*, *29*(5), 385–396. https://doi.org/10.2165/00002018-200629050-00003

Home. (n.d.). https://kebs.isolutions.iso.org/obp/ui#!iso:std:iso:9241:-11:ed-2:v1:en

Hyzy, M., Bond, R., Mulvenna, M., Bai, L., Dix, A., Leigh, S., & Hunt, S. (2022). System Usability Scale Benchmarking for Digital Health Apps: Meta-analysis. JMIR mHealth and uHealth, 10(8), e37290. https://doi.org/10.2196/37290

Inch, J., Watson, M. C., & Anakwe-Umeh, S. (2012). Patient versus healthcare professional spon- taneous adverse drug reaction reporting: A systematic review. https://doi.org/10.2165/ 11631650-000000000-00000

Kiguba, R., Zakumumpa, H., Ndagije, H. B., Mwebaza, N., Ssenyonga, R., Tregunno, P., Harri- son, K., & Pirmohamed, M. (2023). Facilitators and Barriers to Uptake of the Med Safety Mobile App for Adverse Drug Reaction Reporting by Health Workers in Uganda: A Qual- itative Study. *Drug Safety*, *46*(6), 565–574. https://doi.org/10.1007/s40264-023-01303-6

Lewis, D. J., & McCallum, J. F. (2020). Utilizing Advanced Technologies to Augment Pharma- covigilance Systems: Challenges and Opportunities. *Therapeutic Innovation and Regula- tory Science*, *54*(4), 888–899. https://doi.org/10.1007/s43441-019-00023-3

Li, Y., Liang, S., Zhu, B., Liu, X., Li, J., Chen, D., Qin, J., & Bressington, D. (2023). Feasibil- ity and effectiveness of artificial intelligence-driven conversational agents in healthcare interventions: A systematic review of randomized controlled trials. *International Journal of Nursing Studies*, *143*, 104494. https://doi.org/10.1016/j.ijnurstu.2023.104494

Liu, X., Eshghi, A., Swietojanski, P., & Rieser, V. (2021). *Benchmarking Natural Language Un- derstanding Services for Building Conversational Agents* (Vol. 714). Springer Singapore. https://doi.org/10.1007/978-981-15-9323-9_15

M,J. (Ed.). (2023, September 8). Who we serve. Woebot Health. https://woebothealth.com/who-we-serve/

Mittal, M., Battineni, G., Singh, D., Nagarwal, T., & Yadav, P. (2021). Web-based chatbot for Frequently Asked Queries (FAQ) in Hospitals. *Journal of Taibah University Medical Sci- ences*, *16*(5), 740–746. https://doi.org/10.1016/j.jtumed.2021.06.002

Palaian, S., Alshakka, M., & Mohamed Izham. (2010). Developing a consumer reporting

program in Malaysia: A novel initiative to improve pharmacovigilance. *Pharmacy World and Sci- ence*, *32*(1), 2–6. https://doi.org/10.1007/s11096-009-9342-8

Pappa, D., & Stergioulas, L. K. (2019). Harnessing social media data for pharmacovigilance: a review of current state of the art, challenges and future directions. *International Journal of Data Science and Analytics*, *8*(2), 113–135. https://doi.org/10.1007/S41060- 019- 00175-3

Podrazhansky, A., Zhang, H., Han, M., & He, S. (2020). A chatbot-based mobile application to predict and early-prevent human mental illness. *ACMSE 2020 - Proceedings of the 2020 ACM Southeast Conference*, 311–312. https://doi.org/10.1145/3374135.3385319

Sales, I., Aljadhey, H., Albogami, Y., & Mahmoud, M. A. (2017). Public awareness and perception toward Adverse Drug Reactions reporting in Riyadh, Saudi Arabia. *Saudi Pharmaceuti- cal Journal*, *25*(6), 868–872. https://doi.org/10.1016/j.jsps.2017.01.004

Tase, A., Buckle, P., Ni, M. Z., & Hanna, G. B. (2021). Medical device error and failure reporting: Learning from the car industry. *Journal of Patient Safety and Risk Management*, *26*(3), 135–141. https://doi.org/10.1177/25160435211008273

Tase, A., Ni, M. Z., Buckle, P. W., & Hanna, G. B. (2022). Current status of medical device malfunction reporting: using end user experience to identify current problems. *BMJ open quality*, *11*(2). https://doi.org/10.1136/bmjoq-2022-001849

Turk, E., Wontor, V., Vera-Muñoz, C., Comnes, L., Rodrigues, N., Ferrari, G., & Moen, A. (2022). Human-centered integrated care pathways for co-creating a digital, user-centric health information solution. *Journal of Integrated Care*, *30*(4), 296–309. https://doi.org/10. 1108/JICA-01-2022-0007

Van Grootheest, K., De Graaf, L., & De Jong-van Den Berg, L. T. (2003). Consumer adverse drug reaction reporting: A new step in pharmacovigilance? https://doi.org/10.2165/00002018- 200326040-00001

van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies 33-36. *Nursing Stan- dard.*, *16*(40), 33–36. www.nursing-standard.co.uk

Vasquez, G., Bendell, R., Talone, A., Nguyen, B., & Jentsch, F. (2018). The use of immersive virtual reality for the test and evaluation of interactions with simulated agents. Advances in Intelligent Systems and Computing, 15–25. https://doi.org/10.1007/978-3-319-94223-0_2

Appendix A

Imagine you are Sarah Johnson, a 27-year-old female. You were diagnosed with diabetes type 1 when you were 11. Around six months ago you started using a continuous glucose monitoring system to track the levels of your blood sugar. The system works in the following way:

- Once every ten days you attach a sensor via adhesive to your upper arm. This sensor needs to be changed every ten days to provide accurate measures.

- You attach a transmitter to the sensor. This transmitter can be attached to multiple sensors but needs to be changed with a new one every 90 days.

- You connect your sensor to a phone app and every 5 minutes it updates you on your blood sugar levels.

You have a subscription plan with the company Dexcom Inc., and you are delivered a new package every 90 days with the needed supplies for the period until your next package. So far, you have not experienced any problems using the system and you are very satisfied with it.

However, after you changed your last sensor, you started experiencing issues with the Bluetooth connection to the app. On the 2nd of April this year you noticed that your device disconnected unprompted from the app on your phone. Luckily this happened when you were feeling okay and you weren't checking your blood sugar levels to administer insulin, therefore you did not need the results it provides immediately. You checked your phone, and the Bluetooth was on so there was seemingly no reason for the issue. The connection was restored at some point, but this was also very unexpected, and you don't have any explanation for what fixed the issue.

Since then, this has happened several more times, luckily all of them when you were home and if you needed to check your glucose levels you used your old device with testing strips. This was the last sensor of your package, and you used your transmitter until the end of the 90-day period and after that you threw it away. You had your new package with supplies delivered on the 11th of April. The new transmitter has had no issues so far and you have changed several sensors with it already.

Right now, you continue using the Dexcom system since it still provides accurate data, which you tested by comparing it to the results from your old device with test strips. Your healthcare provider advised you to submit a report for an official investigation to the national health agency

since you might have been given a faulty transmitter in your last package. You asked what information you might possibly have to provide, and they advised you to take note of when the incident happened, what went down and what were the effects, as well as to have any relevant information about the device and about yourself prepared since it might be required to submit a report. Luckily, you still had the box of your device at home:


Dexcom ONE
Dexcom ONE CGM System
REF STP-DO-102
LOT 5295257
SN 522DS4
16-08-2021
05-03-2023
Dexcom, Inc.
San Diego, CA, USA
EC REP
MDSS GmbH
Schiffgraben 41
30175 Hannover
Germany
2797
IP28
STERILE R
2°C
30°C
LBL019543 Rev 002
© 2020 Dexcom, Inc.

Appendix B

This appendix includes a link to a repository with the code used to create and train a chatbot system for the experiment. For this to be accessed, the folder needs to be downloaded and then imported into the Google product ’Dialogflow’. The system allows the user to import the zip folder and make adjustments as desired. The link to the repository can be found here:

https://drive.google.com/drive/folders/1nsNaXeN4f8BUuoOzR05l4NcaSKOZtSbI?usp=sharing

Appendix C

Welcome to the reporting form

Here you can report suspected side effects to medicines, vaccines, e-cigarettes, medical device incidents, defective or falsified (fake) products to ensure safe and effective use.

This section will collect information on your topic of reporting. Please provide as much information as possible

What type of medical product do you want to report about? (required) 0 Medical Device 0 Medicine 0 Vaccine 0 Other (please indicate what are you reporting)

Enter the name of the product you want to report (required)

Name of manufacturer/supplier (optional)

Device serial number (SN) (optional)

Batch number/LOT number as indicated on the packaging (optional)

Where did you get the device from? (optional) 0 NHS Hospital 0 Pharmacy 0 Nurse 0 NHS Clinic 0 Private Hospital 0 Shop 0 GP Surgery 0 Private Clinic 0 Mail order/ Internet 0 Other

Manufactured date of the medical device if available (optional) DD MM YYYY

Expiry date of the medical device (optional)

DD MM YYYY

Where is the device currently? (required)

This section will collect information on the incident with the medical device. Please provide as much information as possible

Incident date (optional): DD MM YYYY

Please let us know how severely you, or the patient you are reporting about, were affected by the reaction(s) (optional):

Please describe what went wrong with the device including faults with the device or harm experienced. For example were you unable to obtain a sample, did something break etc. We need this information in order to investigate this incident report. (required)

Please do not add identifiable patient information to the free text.

This section will collect information about you as the person who is reporting an incident as well as ask about the person who was involved in the incident.

I am a ..... (required)

Are you reporting on behalf of yourself or somebody else? (required)

Please provide at least one of the following details about the patient affected by this incident - initials, sex or age. (required)

Appendix D

This appendix contains the scoring table for the form reporting system (FRS). It includes identifiers for all of the mandatory information, as well as what points each question is assigned. The final score is calculated as a percentage of the points obtained out of the total points given.

Question	Required	Correct answer in the form
Type of product	YES	1
Product name	YES	Dexcom; Dexcom One
Adress of manufacturer		-
Name of manufacturer supplier		Dexcom, Inc.
Device SN		522DS4
Batch number/LOT number		5295257
Obtained via		(9) Mail order/Internet
Date of manufacturing		16-08-2021
expiry date		05-03-2023
Current state	YES	(2) Patient/User ; (6) Discarded
Incident date		02-04-2023
Severity		(1) Not erious; (2) Mild to slightly uncomfortable; (3) Uncomfortable, a nuisance or irritation, but able to carry on with everyday activities
Details		***
Reporting person	YES	(1) Member of the public
Affected person	YES	(1) Myself
Patient information	YES	Name - Sarah Johnson
		Sex - Female
		Age - 27

Total score
Minimum required score

Answer given	Scoring in form
	1
	1
	-
	1
	1
	1
	1
	1
	1
	1
	1
	1 - 5
	0.5
	0.5
	1 - 3
20	**12**
*all required questions MUST be filled in, regardless of correction	

Appendix E

This appendix contains the scoring table for the chatbot reporting system (CRS). It includes identifiers for all of the mandatory information, as well as what points each question is assigned. The final score is calculated as a percentage of the points obtained out of the total points given.

Question	Required	Correct answer in the form
Type of product	YES	1
Product name	YES	Dexcom; Dexcom One
Adress of manufacturer		-
Name of manufacturer supplier		Dexcom, Inc.
Device SN		522DS4
Batch number/LOT number		5295257
Obtained via		(9) Mail order/Internet
Date of manufacturing		16-08-2021
expiry date		05-03-2023
Current state	YES	(2) Patient/User ; (6) Discarded
Incident date		02-04-2023
Severity		(1) Not erious; (2) Mild to slightly uncomfortable; (3) Uncomfortable, a nuisance or irritation, but able to carry on with everyday activities
Details		***
Reporting person	YES	(1) Member of the public
Affected person	YES	(1) Myself
Patient information	YES	Name - Sarah Johnson

Answers accepted by bot	Versions of the answer
medical device; a device; blood sugar monitoring device; transmiter;blood sugar monitor; dexcom device	
Dexcom; Dexcom One	
Dexcom,Inc. San Diego, CA, USA	
-	
522DS4	
5295257	
-	
16/08/2021; 16-08-2021	
05/03/2023; 05-03-2023	
Stopped using it; Changed transmitters; Discarded the faulty device; Threw it away	
02/04/2023; 02-04-2023	
Not harmed; No harm; No lasting consequences; Not severe	
***	
A regular person; A regular user;	
To me; It happened to me; I was involved in the incident; Me	
Name - Sarah Johnson	At least one of the information units should be given, but it is not reuqired to fill in all three. Valid anse can also be any combination between two information points.
Sex - Female/Woman Age - 27	
	Total score
	Minimum required score

Scoring in form	Scoring in chat-bot log	
	1 1	
		1 1
- 1		
	1 -	
		1 1
	1 1	
-	-	
		1 1
		1 1
	1 1	
		1 1
	1 1	
		0-5 0-5

1 1	
1 1	
patially answered the question - 0.5 1	0.5 -
fully answered the question - 1.0	
19	**19**
*all required questions MUST be filled in, regardless of correction	*report needs to be finished

Appendix F

This appendix contains the answers given to the question asking for detailed descriptions in own words of the incident. The answers are colour-coded in accordance with what piece of information matches which information requirement, as listed at the end of the document, and it corresponds to the number of points given for the response.

exerpts from used answers

The device disconnects at random times; (the device disconnected at random times from my phones bluetooth)

The transmitter disconnected from the bluetooth several times without any clear reason

It keeps disconnecting from the app on my phone

The device kept disconnecting and conecting to my phone
point the Bluetooth connection was disrupted and I could not get any data on y phone app
the phone
Device disconnected from the app
from the bluetooth app
the phone
noticed that my device disconnected unprompted from the app on my phone. Luckily this happened when I was feeling okay and weren't checking my blood sugar levels to administer insulin, therefore I did not need the results it provides immediately. I checked my phone, and the Bluetooth was on so there was seemingly no reason for the issue. The connection was restored at some point, but this was also very unexpected, and I
disconnects with the app on the phone, issues with bluetooth connection
The device was disconnected from Bluetooth several times without notifying and there were no error messages as to explain why this occured.
No data on mobile app
problem with connection with phone apps
The device disconnected although bluetooth was still on
connection between the app on my phone and my transmitter was interrupted several times. This meant I could not check my glucose levels using the transmitter; I had to use an older device with testing strips. The connection seemed to randomly restore itself
device disconnects unprompted from the phone app
phone was enabled.
The bluetooth connection to the app from the device started disconnecting and as a consequence the blood levels were not immediatly available. The issue seemed to occur for no reason. The connection was restored at some point but the issue seemed to repeat itself. After gettting a new device the issue has no happened so far.

phone. It disconnected a few times even though I had my bluetooth on and I could not pinpoint the problem. It reconnected eventually, but I still don't know what happened. Luckily, it never disconnected when it was a serious time.
On the 2nd of April I noticed the monitor stopped giving information to the app on my phone. It seemed to had been disconnected. This happened a couple other times after. on 02/04/2023, the device disconnected from the app on my phone. It was lucky I wasn't checking my levels at this time and I was feeling well. My bluetooth was on and I had no other problematic connections. This resolved, but also happened several times over the 90 day period. This is a real concern for managing my health. Luckily on these occasions I was feeling well and my levels were controlled. Since then, my new order has been

The device disconnected from my phone despite my Bluetooth being on. After some time it connected again and remained stable. I've had this issue several times.

I started experiencing issues with the Bluetooth connection of this device. This occurrence already happened a few times…
It arrived broken and had some severe side affects in regard to skin
The device unexpectedly and unexplainably disconnected and reconnected from the app on my phone several times even though the bluetooth on my phone was on.
It kept disconnecting with my phone on several occasions
April 2023.
randomly.
sugar levels, which endangers the patient and causes distress.
case there were no serious consequences there could have been serious consequences due to the connection issues.

Appendix G

This appendix contains a link to a folder with all additional material used during work on this research. This includes:

- Advertising materials for participant recruitment
- Consent form given to participants
- The original data file obtained from data collection
- Edited data after applying the exclusion criteria
- Explicit scoring per participant for report completeness
- R script to reproduce the data analysis. Disclaimer: This script was written to work with the R and R studio version (2023.06.2 Build 561).

The link to access all of the materials will be active until January 2026.

https://drive.google.com/drive/folders/1P02jXjQnPmlW7-BVb46exD7ql1QzVjkO?usp=drive*ink*