



UNIVERSIDAD PERUANA
CAYETANO HEREDIA

“EVALUACIÓN DEL USO DE UN
CHATBOT
PARA EL SEGUIMIENTO EN UN
ENSAYO
CLÍNICO DE PROFILAXIS FRENTE AL
COVID-19 EN PERSONAL DE SALUD”

TESIS PARA OPTAR EL GRADO DE
MAESTRO EN INFORMÁTICA
BIOMÉDICA EN SALUD GLOBAL CON
MENCIÓN EN INFORMÁTICA EN
SALUD

ANTHONY BACILIO RUIZ

LIMA-PERÚ

2021

ASESORA:

Dra. Patricia Jannet García Funegra

JURADO DE TESIS

DR. JESUS LORENZO CHIRINOS CÁCERES

PRESIDENTE

DR. GERMAN FELIPE ALVARADO CUTIPA FLORES

VOCAL

MG. DANIEL FLAVIO CONDOR CAMARA

SECRETARIO (A)

DEDICATORIA

A mis padres, mis hermanos, mi mamita y mis familiares, por apoyarme mi decisión de seguir este camino de la investigación a pesar de todas las desventajas que este presenta y brindarme soporte emocional y psicológico durante toda mi vida y más estos dos últimos años.

También a mi abuelo, porque pudo guiar el camino de su familia con su terquedad y dedicación. Te recordaremos por siempre.

AGRADECIMIENTO

A CONCYTEC por haberme financiado y permitido realizar la Maestría en
Informática Biomédica en Salud Global.

A la Dra. Patricia García por su tiempo y paciencia por iluminar mi mente y mostrarme como se debe hacer investigación correctamente y seguir apoyándome en medio de esta guerra contra el SARS-Cov-2 y al Dr. Cesar Carcamo por iluminarme en el camino del análisis de datos.

Agradezco al PhD. Alejandro Llanos y a los investigadores pertenecientes al ensayo clínico “Hidroxicloroquina para prevenir infección por SARS-CoV-2 en personal de salud: Ensayo clínico controlado de fase 3, aleatorizado, de etiqueta abierta” por permitirme evaluar al chatbot de seguimiento.

A los miembros del equipo de monitoreo, Álvaro Schwalb, Juan Luis Quintana, Paulo Aguirre, Fátima Ávila y Mario Damián por su tiempo y apoyo a esta evaluación.

A Paola Pflucker y al equipo de esta Maestría por siempre estar atentos de mí y de los alumnos para que no desistamos en nuestras investigaciones. Así como a Ailin Cabrera y todo el equipo y docentes de la FASPA por las clases recibidas, las preguntas absueltas y el tiempo que hemos podido compartir.

También agradezco a mis compañeros de la Maestría por el apoyo brindado.

TABLA DE CONTENIDOS

| | |
|---|----|
| 1. Introducción | 1 |
| 1.1. Antecedentes | 1 |
| 1.2. Planteamiento del Problema..... | 3 |
| 1.3. Marco Teórico..... | 4 |
| 1.3.1. El SARS-CoV-2..... | 4 |
| 1.3.2. Hidroxicloroquina (HCQ) | 6 |
| 1.3.3. “Hidroxicloroquina para prevenir infección por SARS-CoV-2 en personal de salud: Ensayo clínico controlado de fase 3, aleatorizado, de etiqueta abierta”. | 7 |
| 1.3.4. Inteligencia Artificial (IA) | 8 |
| 1.3.5. Chatbots | 12 |
| 1.3.6. Evaluación de un servicio de chatbot..... | 14 |
| 1.3.7. Evaluaciones de usabilidad. | 17 |
| 1.3.8. Chatbots en diversos ámbitos..... | 18 |
| 1.3.9. Customer Relationship Management (CRM)..... | 19 |
| 1.4. Justificación de la Investigación | 20 |
| 2. Objetivos | 22 |
| 2.1. Objetivo general..... | 22 |
| 2.2. Objetivos secundarios | 22 |
| 3. Metodología | 23 |
| 3.1. Diseño del Estudio | 23 |

| | |
|---|----|
| 3.2. Población..... | 24 |
| 3.3. Comunicación con el Chatbot | 24 |
| 3.3.1. Funcionamiento diario del sistema chatbot..... | 24 |
| 3.4. Instrumentos usados | 26 |
| 3.4.1. Registro de Participantes..... | 26 |
| 3.4.2. Escala de Usabilidad del Sistema..... | 26 |
| 3.4.3. Cuestionario de utilidad del Sistema para el equipo de investigación | 29 |
| 3.5. Procedimientos..... | 32 |
| 3.5.1 Diseño del Sistema..... | 32 |
| 3.5.2 Piloto del Sistema..... | 32 |
| 3.5.3 Construcción en el sistema de “Customer Relationship Management” | 32 |
| 3.5.4 Entrenamiento para su uso | 33 |
| 3.5.5. Implementación..... | 33 |
| 3.5.6. Evaluación..... | 34 |
| 3.6. Análisis de la data | 38 |
| 3.7. Consideraciones éticas | 39 |
| 4. Resultados | 40 |
| 4.1 Características demográficas de los participantes del ensayo..... | 40 |
| 4.2. Métricas de usuarios y del chatbot | 41 |
| 4.3. Evaluación de usabilidad por parte de los participantes | 43 |

| | |
|--|----|
| 4.4. Usabilidad y utilidad por parte del equipo a cargo de la operación del sistema (monitores) | 45 |
| 4.4.1. Evaluación de Usabilidad..... | 45 |
| 4.4.2. Evaluación de utilidad..... | 46 |
| 5. Discusión..... | 50 |
| 6. Conclusiones | 58 |
| 7. Referencias bibliográficas | 59 |
| ANEXO I : Flujo de diálogo | |

TABLA DE FIGURAS

| | | |
|--------|--|----|
| Fig.1 | Perceptrón | 10 |
| Fig.2. | Perceptrón multicapa | 10 |
| Fig.3. | Red Neuronal con LSTM | 11 |
| Fig.4. | Partes de un Chatbot | 13 |
| Fig.5. | Modos de mantener un diálogo | 14 |
| Fig.6. | Plan de implementación y uso del sistema | 33 |

TABLA DE TABLAS

| | | |
|----------|---|----|
| Tabla 1 | Escala EUS e Interpretaciones | 28 |
| Tabla 2 | Características demográficas de los participantes | 40 |
| Tabla 3 | Métricas de las interacciones del bot | 41 |
| Tabla 4 | Métricas del tiempo usado por participante | 42 |
| Tabla 5 | Porcentaje de participantes que están de acuerdo con cada enunciado | 43 |
| Tabla 6 | Evaluación del Puntaje EUS como variable dependiente | 44 |
| Tabla 7 | Porcentaje de Monitores que están de acuerdo con los enunciados de usabilidad | 45 |
| Tabla 8. | Resultados de evaluación de utilidad (Porcentaje de acuerdo con los enunciados) | 46 |

Resumen

Introducción: Actualmente existen sistemas Chatbot que ejecutan conversaciones con múltiples usuarios en simultáneo. Desafortunadamente son pocas las investigaciones sobre uso para problemas de salud o en Investigación.

Objetivos: Evaluar la usabilidad de un Chatbot para el seguimiento de adherencia al tratamiento, y efectos adversos en un ensayo clínico en el Perú.

Métodos: Es un estudio de usabilidad. Se implementó el chatbot y al finalizar el estudio se invitó a los participantes y al equipo de monitoreo a evaluar el sistema con la “Escala de Usabilidad del Sistema.” También, se colectaron métricas del uso del bot.

Resultados: 40 participantes evaluaron el sistema. Se realizaron 3526 interacciones (86.6% fueron satisfactorias) y se ahorraron 2219.82 horas de trabajo en seguimiento y colección de datos gracias al envío y registro automático de datos. El puntaje promedio de usabilidad fue de 78.25 (“Aceptable”), y no se encontraron diferencias significativas según características sociodemográficas. Los monitores calificaron el sistema también con un 71.7 (“Aceptable”).

Conclusión: Este tipo de aplicaciones (chatbots) son muy bien recibidas y podrían adecuarse y usarse más frecuentemente en investigación y también en otras aplicaciones de salud que requieran seguimiento de síntomas y adherencia al tratamiento, como enfermedades crónicas y otras.

Palabras clave: Chatbot, Cumplimiento y Adherencia al Tratamiento, Estudios de Seguimiento, usabilidad. (DeCS/BIREME)

Abstract

Introduction: Nowadays there are chatbot systems that execute conversations with multiple user on simultaneous. Unfortunately, there is small research about the use on health sector and research.

Objective: Evaluate the usability of a Chatbot in the monitoring of adherence to a treatment and adverse effects in a clinical trial in Peru.

Methods: This is a usability study. The chatbot was implemented and at the end of the study, the participants and the monitoring team were invited to evaluate the system with “System Usability Scale”. Also, chatbot metrics were collected.

Results: 40 participants evaluated the system. 3526 interactions were taken (86.6% were successful) and 2219.82 hours of work were saved on monitoring and data collection thanks to the automated data send and registration. The average usability score was 78.25 (“Acceptable”) and there were no significant differences from social-demographic characteristics. The monitors rate the system with a 71.7 (“Acceptable”)

Conclusion: This kind of systems(chatbots) are welcomed and could be tailored and be used more frequently on research and another health sector tasks that require symptoms and adherence monitoring, like chronic and other diseases.

Keywords: Chatbot, Treatment Adherence and Compliance, Follow-Up Studies, usability (MeSH/NLM)

Lista de Abreviaturas

| | |
|------------|---|
| API | Interfaz de programación de Aplicaciones |
| COVID-19 | Enfermedad por Coronavirus 2019 |
| CQ | Cloroquina |
| CRM | Client Relationship Management (Manejo de la relación con los clientes) |
| CSUQ | Computer Systems Usability Questionnaire (Cuestionario de Usabilidad de Sistemas) |
| DE | Desviación Estándar. |
| EC | Ensayo Clínico |
| EPP | Equipos de Protección Personal |
| EUS | Escala de Usabilidad del Sistema |
| HCH | Hospital Cayetano Heredia |
| HCQ | Hidroxiclороquina |
| HHU | Hospital Hipólito Unanue |
| HNAL | Hospital Nacional Arzobispo Loayza |
| IA | Inteligencia Artificial |
| INS | Instituto Nacional de Salud |
| LSTM | Long Short Term Memory |
| RC | Reflection Companion |
| RD | Resolución directoral |
| RNA | Ácido Ribonucleico |
| SARS-CoV-2 | Coronavirus de del Síndrome Respiratorio Agudo Severo 2 |
| SIDISI | Sistema Descentralizado de Información y Seguimiento a la Investigación |
| SUS | System Usability Scale |
| TE | Tiempo de espera |
| TI | Tiempo de interacción |
| TIC | Tecnología de la Información y computación |
| TT | Tiempo total de uso de sistema por participante |
| UCI | Unidad de Cuidados Intensivos |
| UMUX | User Metric for Usability Experience (Métrica de Usuario para la Experiencia del Usuario) |

1. Introducción

1.1. Antecedentes

El COVID-19 causado por el coronavirus SARS-CoV-2 ha ocasionado la pandemia que actualmente se vive en el mundo. Este se inició en la ciudad de Wuhan en la República Democrática de China [1] y ha causado más de 153 millones de infecciones y más de tres millones de muertes alrededor del mundo [2]. Asimismo debido a la alta transmisibilidad del virus sumado a la sobrexposición del personal de salud a la infección y la escasez de Equipos de Protección Personal (EPP), ha ocasionado más de 13 mil casos de infección, 35 ingresos a Unidad de Cuidados Intensivos (UCI) y 450 decesos en el personal médico nacional[3].

A inicios de mayo del 2020, no existían vacunas que permitan prevenir la infección ni medicamentos que sean efectivos contra el Virus. Para reducir los tiempos para una respuesta sobre la efectividad de los medicamentos, se usó la estrategia de investigar drogas que ya han sido usadas para otras enfermedades y que tengan actividad, al menos in vitro, contra el SARS-CoV-2. Una de estas fue la Hidroxicloroquina(HCQ) [4]. Ya hay, al menos, 266 ensayos clínicos evaluando la HCQ [5] como profilaxis o tratamiento. La gran mayoría con números pequeños de casos. Pero al menos dos mega ensayos, RECOVERY [6,7] y SOLIDARITY[8,9], han demostrado la futilidad (inutilidad) de la HCQ para el manejo de pacientes hospitalizados (moderados y severos). El Perú es uno de los países que está participando en SOLIDARITY.

Sin embargo, aún no quedaba claro si la HCQ poseía un rol en el manejo de los pacientes leves o si podría ser efectiva como profilaxis de la infección, especialmente en poblaciones con alta exposición, como el personal de salud. En el mundo se han registrado 69 estudios de profilaxis con HCQ para el COVID-19 [5]. Uno de estos estudios fue realizado por un grupo de investigadores peruanos, quienes diseñaron un ensayo aleatorizado para evaluar HCQ en personal de salud en el Perú (“Hidroxicloroquina para prevenir infección por SARS-CoV-2 en personal de salud: Ensayo clínico controlado de fase 3, aleatorizado, de etiqueta abierta” con número de SIDISI 202087, aprobación por INS para el EC No. 011-20 y registro en Clinical Trials NCT04414241) [10]. Este ensayo recibió la aprobación Instituto Nacional de Salud y fue ejecutado entre los meses de junio a noviembre del 2020.

Como parte de los procedimientos del ensayo clínico fue crítico asegurar el seguimiento de los participantes, tanto los que fueran aleatorizados a la intervención como los controles. Se necesitaba registrar si los participantes cumplían con ingerir el medicamento (HCQ) interdiariamente, así como la presencia de síntomas y posibles adversos. Pero realizar esta tarea podría ser complicada y costosa por lo que se planteó usar tecnología para su implementación.

Actualmente existen sistemas de Manejo de relación con Clientes (CRM) y Bots conversacionales (chatbots) que a través de las Tecnologías de la Información y Computación (TICs) y la Inteligencia Artificial (I.A.) pueden ejecutar tareas repetitivas con múltiples usuarios a la vez; entre ellas, llevar una conversación

significativa con una persona [11]. Sin embargo, son pocas las investigaciones sobre el uso de este tipo de tecnologías para el soporte en problemas de salud.

El ensayo aleatorizado de profilaxis con HCQ incluyó el diseño y la implementación del chatbot y estos fueron aprobados por el Comité Nacional Transitorio de Ética del Instituto Nacional de Salud. Para esta tesis planteamos la evaluación de usabilidad del sistema de chatbot para el seguimiento de los participantes del Ensayo Clínico “Hidroxicloroquina para prevenir infección por SARS-CoV-2 en personal de salud [10].

1.2. Planteamiento del Problema

El coronavirus 2 del Síndrome Agudo Respiratorio Grave (SARS-CoV-2) ha originado una pandemia con más de 153 millones de infectados y 3.2 millones de muertes[2]. En el Perú se reportan al menos 1.81 millones de infectados y 62 mil muertes[12].

El personal de la salud representa la primera línea de defensa frente a la pandemia. Esto implica que su riesgo de contagio es muy alto dada su exposición al atender a las personas infectadas. Es así que se hace prioritario buscar estrategias que puedan reducir su riesgo. Con el objetivo de evaluar una posible intervención para prevenir la infección del personal de salud, se realizó un ensayo clínico para la profilaxis con HCQ en el Perú. En este contexto, para hacer más eficiente el trabajo de seguimiento de los participantes dentro del ensayo y reducir las interacciones de personas que podría condicionar riesgos de infección, se propuso el uso de

tecnologías de información, que incluyan servicios automatizados con mensajería instantánea e integración con chatbots.

Con el respaldo de la tecnología y frente a la necesidad de hacer seguimiento a los participantes en este ensayo, se consideró una buena oportunidad realizar el diseño, implementación y evaluación de este tipo de solución tecnológica para salud en el contexto de la pandemia del COVID-19.

1.3. Marco Teórico

El siguiente marco teórico explicará los conceptos necesarios para entender el trabajo de investigación. Se comenzará brevemente por una explicación de que es el SARS-CoV-2, su importancia y las actuales medidas para prevenirlo. Luego se continuará con el uso de la Hidroxicloroquina, su importancia como posible uso de profilaxis, porque es necesario el seguimiento y una descripción breve del ensayo. Más adelante, se explicará sobre la inteligencia artificial y el desarrollo de los Chatbots, cómo se evalúan y cómo se usan en un CRM para el manejo de clientes.

1.3.1. El SARS-CoV-2

El Coronavirus 2 del Síndrome Respiratorio Agudo Grave (SARS-CoV-2) es un virus de la familia de los coronavirus[13]. Estos tienen un solo filamento de RNA y sus diferentes mutaciones causan enfermedades en diferentes especies, entre ellas, la humana [14]. Para infectar a las células humanas utilizan el receptor ACE2 de las células, el cual posee las características necesarias para unirse con las

glicoproteínas, las cuales dan energía, pero también son parte de la estructura de este coronavirus [15].

Este virus es causante de la enfermedad denominada COVID-19, cuyos síntomas son fiebre, tos, disnea y puede llegar hasta la neumonía, requiriendo cuidados intensivos y poniendo en peligro la vida de las personas [16]. Además, dada su alta transmisibilidad y lo interconectado que se encuentra el mundo, en un tiempo relativamente corto, se ha convertido en una pandemia[17] [18].

El COVID-19 ha trastocado al mundo en general, desnudando las debilidades de los sistemas de salud y ha condicionado una sobre exposición del personal sanitario, que representa uno de los grupos altamente afectados[19].

Para prevenir el contagio en la población general se ha recomendado lavarse las manos frecuentemente, el uso de mascarilla, ventilación y el distanciamiento social [20].

Para el personal de salud, la protección se basa en el seguimiento de las reglas de bioseguridad y el uso de Equipo de Protección Personal (EPP) [21]. Sin embargo, el personal médico aún sigue infectándose debido a la escasez de EPP's, su uso incorrecto, la pobre ventilación en los establecimientos de salud y la continua sobreexposición [22].

Frente esta situación, fue necesario buscar alternativas adicionales que protejan al personal de la salud.

1.3.2. Hidroxicloroquina (HCQ)

La Cloroquina (CQ) se encuentra naturalmente en el árbol nacional de la Quina, del cual en 1633, en Europa se usaba la corteza para combatir los escalofríos[23]. Aunque, en realidad, era este árbol el que causaba los síntomas [24]. En 1934, Hans Andersag redescubre el compuesto y se empieza a usar como profilaxis y tratamiento antimalárico [25]. La hidroxicloroquina (HCQ) es la forma sintética de la CQ y tiene menores efectos secundarios.

Al inicio de la pandemia se hicieron públicos varios estudios sobre la efectividad in-vitro de la HCQ en la inhibición de los coronavirus [26]. Esto condicionó, en algunos países, su uso sin real data clínica, pero también condicionó que se iniciaran varios estudios para evaluar su efectividad como tratamiento para el COVID-19 [27,28].

Una de las alternativas del uso de HCQ que se planteó fue como profilaxis frente al COVID-19 para grupos con alta exposición como el personal de salud. Al momento que iniciamos nuestro estudio, se registraban al menos 72 ensayos clínicos de profilaxis con HCQ en el mundo[5,29]. Aquí en el Perú, se desarrolló un ensayo clínico nacional denominado “Hidroxicloroquina para prevenir infección por SARS-CoV-2 en personal de salud, el cual se detallará a continuación.

1.3.3. “Hidroxicloroquina para prevenir infección por SARS-CoV-2 en personal de salud: Ensayo clínico controlado de fase 3, aleatorizado, de etiqueta abierta”.

Este ensayo tuvo como objetivo evaluar la efectividad de la profilaxis interdiaria de HCQ para la prevención comparando con las medidas estándar de protección. Para esto se midió por la seroconversión a las 4 semanas y 8 semanas. También, se evaluó la seguridad y tolerabilidad de la profilaxis con HCQ interdiaria.

Este ensayo clínico aleatorizado, controlado y pragmático contó con dos brazos:

- HCQ más medidas de protección personal estándar proporcionadas en el hospital.
- Medidas de protección personal estándar proporcionadas en el hospital.

El ensayo reclutó personal de salud que estaba trabajando durante el brote COVID-19, ya sea como personal médico, personal de enfermería, personal técnico o auxiliar de enfermería y que tenga una prueba molecular y serológica rápida negativa para SARS-CoV-2. Se excluyeron aquellos con sintomatología compatible con COVID-19, que hayan ingerido en los últimos 30 días HCQ, cloroquina sulfato o azitromicina, o tengan antecedentes de enfermedad cardíaca o historial conocido de síndrome de QT prolongado.

El ensayo reclutó al personal de salud en tres centros hospitalarios: Hospital Cayetano Heredia (HCH), Hospital Nacional Arzobispo Loayza (HNAL) y el Centro Medico Naval “CMST” (HN). Después de ser invitados, los participantes, firmaron un consentimiento informado y fueron aleatorizados. De ser aleatorizado

al brazo de intervención, se les administró HCQ interdiaria por 8 semanas. Si durante el estudio el participante obtenía una prueba molecular/serológica para SARS-CoV-2 positiva, se daba por finalizada su participación en el estudio. También, el participante podría ser retirado del estudio por eventos adversos o por decisión propia.

Fue, además, necesario realizar un seguimiento diario de la adherencia al tratamiento en el brazo de intervención, y el seguimiento de síntomas y efectos adversos en ambos brazos. Para ello se implementó el uso de TICs, específicamente de Inteligencia Artificial y chatbots.

1.3.4. Inteligencia Artificial (IA)

La Inteligencia Artificial (IA) ha sido definida de diversas formas. Sin embargo sería bueno entenderla como “El estudio de cómo hacer que las computadoras hagan algo para lo que aún las personas son mejores” [30]. Con este objetivo, el hombre ha hecho que las computadoras no sólo resuelvan tareas que pueden hacer los humanos, sino también sobrepasarlos [31].

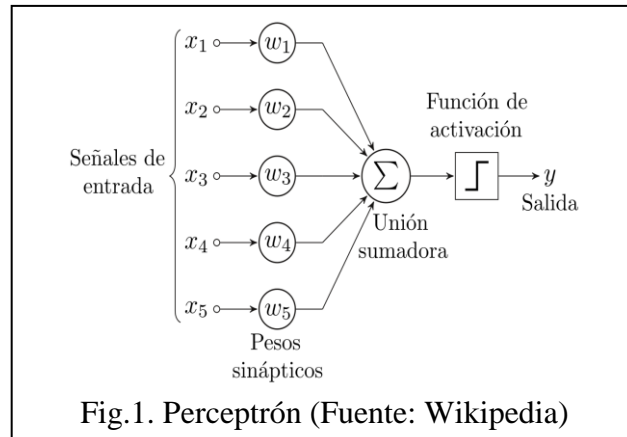
Históricamente, la IA tiene importancia en las ciencias de la computación gracias al trabajo de Alan Turing quien, en 1950, se preguntó si “Las máquinas pueden pensar”[32]. Con esta interrogante en mente, diseñó el Test de Turing, con el cual un usuario “ciego” dialoga textualmente con una máquina y una persona. Luego tiene que identificar cual es cada uno solo a través del diálogo.

Después, en 1958, Frank Rosenblatt creó los perceptrones que fueron inicios de las redes neuronales [33]. En 1964, Daniel Bobrow desarrolló un programa llamado STUDENT para resolver programas de álgebra escrito[34]. En los años 80, se enfocó el estudio en sistemas basados en conocimientos[35]. En los años 90, surgieron los agentes inteligentes que podían analizar el entorno para maximizar las probabilidades de éxito[36]. En 2011, el ordenador Watson de IBM ganó el concurso de preguntas y respuestas “Jeopardy”[37].

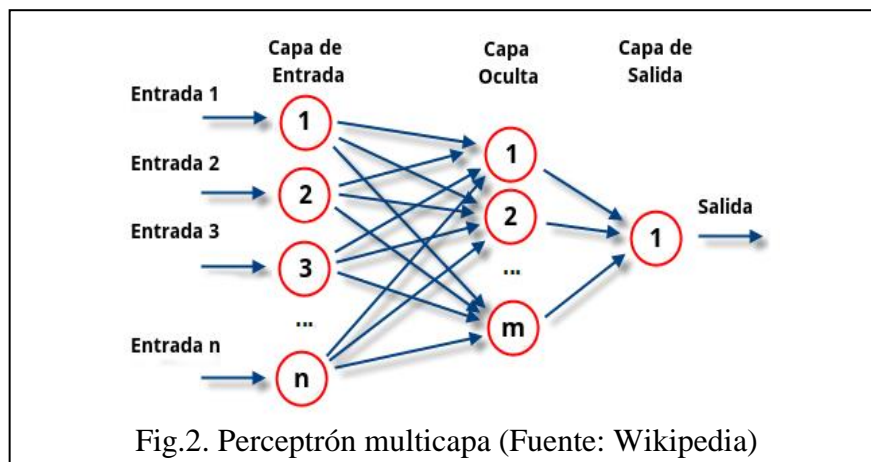
Las computadoras han logrado del mismo modo resolver diferentes tipos de tareas como:

- resolución de rompecabezas [38]
- representación del conocimiento [39]
- planeamiento [39]
- autoaprendizaje [40]
- procesamiento de lenguaje natural (comunicarse con humanos)[41]
- manipulación y movimiento [42]
- Inteligencia social [43]

Una de las herramientas que han revolucionado las ciencias computación para la inteligencia artificial son las redes neuronales artificiales [44]. Estas redes neuronales artificiales están compuestas por neuronas computacionales conocidas como perceptrones [45]. Estos, se pueden entender como unas piezas de programación que tienen nodos de entradas y de salida (Figura 1).

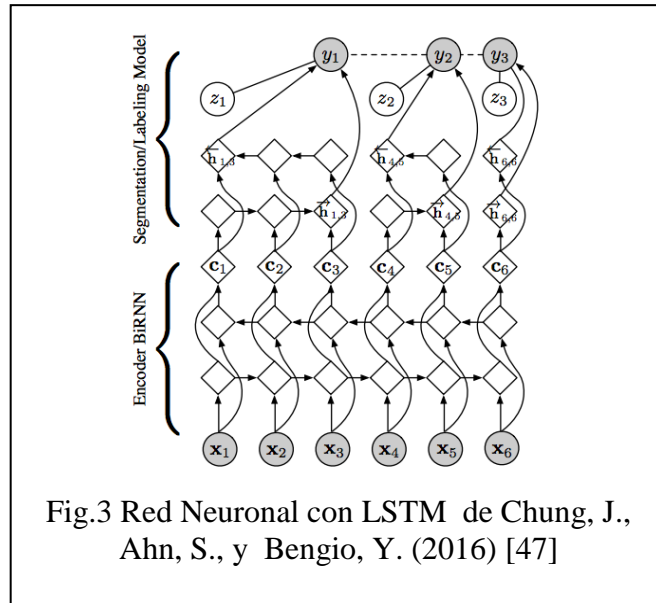


Esta pieza utiliza los valores que ingresan en los nodos de entradas ponderados por unos valores omega (ω_1, ω_2 , etc) y sumados. A través de una función de activación, genera resultados en los nodos de salida. Luego, si se une los nodos de salida como entradas para siguientes neuronas, se formarán las redes neuronales [44] (Figura 2).



Si se siguen conectando perceptrones las redes irán creciendo pudiendo llegar a formar redes inmensamente amplias que se denominan “Redes neuronales profundas” que sirven para lo que se denomina “Aprendizaje profundo”[46]. Ahora, si uno de los nodos de salida de algunas de las neuronas se conectan a nodos de entrada de neuronas anteriores a este, se tendrán “Redes neuronales recurrentes”

[47]. Estas son el corazón de las redes con gran memoria de corto plazo (LSTM acrónimo del inglés Long short term memory) [48] (Figura 3).



Ahora es adecuado explicar cómo estas redes se entrenan para ser inteligentes. El entrenamiento consistirá en ajustar estos valores ω con algoritmos diferentes como corrección por el tamaño de error cometido (*backpropagation*) [49] con el objetivo de que para determinadas entradas me den los valores de salida que deseo. De esta forma, los valores de entrada podrían ser caracteres de un texto y las salidas podrían ser un valor que identifique una acción frente a este texto.

Si bien elaborar un sistema de procesamiento de lenguaje natural desde cero es complicado, existen formas accesibles de usar esta tecnología.

Actualmente con la proliferación de los servicios de mensajería instantánea y la necesidad de fidelizar a los clientes a través de la atención se han creado sistemas que utilizan inteligencia artificial para ofrecer servicios personalizados. Unos son

los llamados servicios de chatbots para mensajería instantánea, los cuales ahorran los tiempos de implementación y los costos de mantenimiento.

1.3.5. Chatbots

Los Chatter bots, chatbots, bots conversacionales, agentes virtuales conversacionales, bots, etc. son programas computacionales que tienen una “personalidad” y pueden ejecutar una conversación significativa con una persona [11].

Los chatbots están estrechamente relacionado con la IA. [32], y últimamente se ha popularizado.

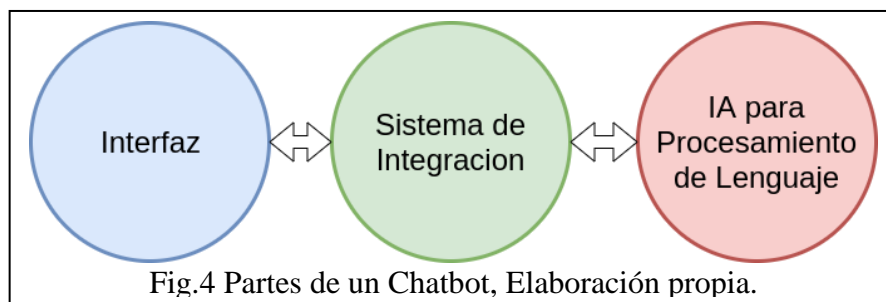
Cabe destacar que los servicios de bot conversacional no son solo IA, ya que necesitan de un interfaz que puede ser los servicios de mensajería social. Por eso, algunos están en la web, otros en redes de mensajería social o aplicativos.

Todo servicio de bot conversacional cuenta con 3 componentes [50] (Fig. 4):

- Primero es el servicio de IA que realizará el procesamiento de lenguaje natural.
 - Ejemplos de esto son los servicios de Dialogflow o Twilio[51].
- Segundo es el servicio de integración que une la IA con la interfaz de usuario.
 - Este usualmente lo provee quien ofrece el servicio de procesamiento de lenguaje natural.

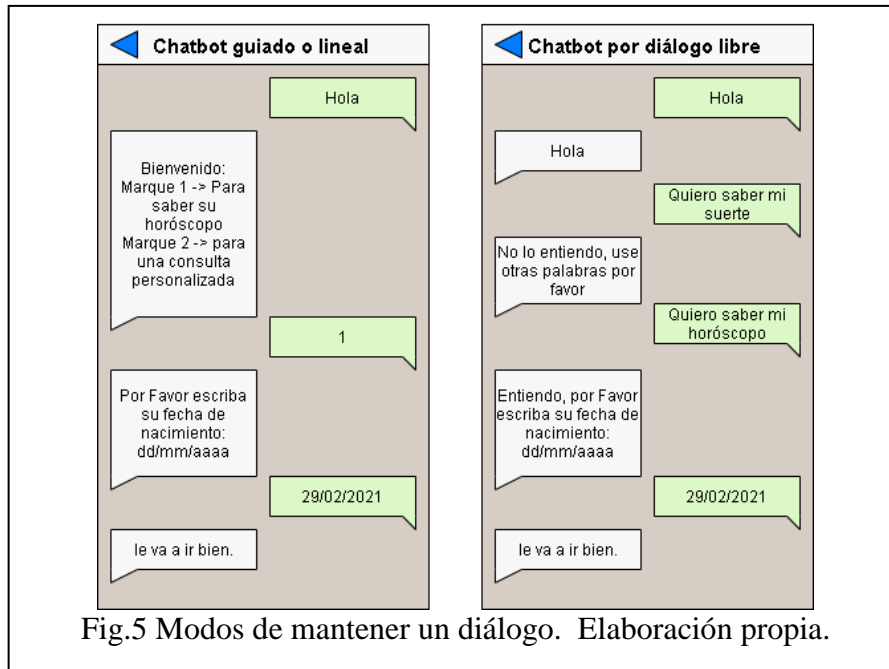
- Tercero es la interfaz que tiene contacto con el usuario.
 - Puede ser una página web, un aplicativo dedicado o un servicio de mensajería social.

Es importante mencionar que no todos los chatbots conversacionales son autosuficientes, sino que algunos se conectan con otros servicios, por ejemplo, aquellos que permiten pedir pizzas o separar vuelos.



Los chatbots poseen diversos modos de mantener un diálogo, pero existen 2 formas comunes de hacerlo:

- Diálogo guiado o lineal a través de opciones (Figura 5.a)
 - Ventaja: El usuario puede saber qué puede conseguir con el servicio.
 - Desventaja: No es tan natural y si hay muchas opciones el usuario se puede sentir sobrecargado de información.
- Diálogo libre a través de una conversación abierta (Figura 5.b)
 - Ventaja: Es más natural
 - Desventaja: El usuario no sabe que puede ofrecerle el chatbot y el chatbot puede no entender ciertas consultas.



1.3.6. Evaluación de un servicio de chatbot

Si bien, actualmente la implementación de un chatbot puede no ser muy complicada, su diseño puede tomar tiempo y dependiendo de los casos, puede ser costoso. Por esta razón, es clave la medición de su usabilidad y efectividad.

Como se mencionó antes, una de las primeras evaluaciones de herramientas que usan IA fue el Test de Turing [52]. Con el fin de ver que tan humana podría ser vista la comunicación de la IA, se pedía a un evaluador “ciego” que se comunicara de forma escrita con una máquina y una persona. Luego, tenía que identificar si el emisor era máquina o era persona.

Con el desarrollo de las herramientas informáticas y las redes neuronales para la IA, se pudo iniciar las evaluaciones del entrenamiento del chatbot. Para esto se

dividían bases de datos de preguntas con su respuesta correspondiente en 2 partes. Se entrenaba el chatbot con la base de datos más grande y se evaluaba la precisión de selección de respuestas con la base de datos pequeña [48].

Después, con el objetivo de implementar servicios que utilicen chatbots, se ha visto necesario evaluar la calidad de versiones del chatbot. Esto se puede hacer, por ejemplo, a través de un Proceso Jerárquico Analítico (Analytic Hierarchy Process) [53]. De este modo, los desarrolladores o las empresas pueden evaluar una variedad de atributos de calidad entre dos versiones de un mismo chatbot. Estas pueden ser:

- Eficiencia
- Funcionamiento:
- Robustez a entradas inesperadas, evasión de declaraciones inapropiadas, etc
- Efectividad
- Funcionalidad: Interpretación, Uso apropiado de la formalidad, etc.
- Humanidad: Test de Turing, Mantener la conversación sobre un tema, etc.
- Satisfacción
- Afecto: Provee saludos adecuados, Lectura y respuesta frente a diferentes estados de ánimo, etc
- Ética y comportamiento: Conocimiento ético y cultural de los usuarios, Protección y respeto de la privacidad, etc.
- Accesibilidad: Puede detectar la intención de la pregunta, etc.

Y para evaluar lo anterior se puede usar diferentes métricas, como:

- Efectividad de la pregunta contestada
- Porcentaje de éxito
- Porcentaje de clasificación correcta.
- Cualquier métrica que pueda servir para medir los atributos de calidad.

Ahora, por ejemplo, si se considera el uso del bot conversacional como parte del servicio de atención al cliente en una empresa comercializadora, se puede medir la Analítica Conversacional del Bot [54] que incluye las siguientes métricas:

Métricas de usuario:

- Total de usuarios: El número de personas que utilizan el chatbot.
- Usuarios activos: Los usuarios que leen tus mensajes antes de un margen de tiempo
- Usuarios enganchados: Usuarios que envían mensajes y leen mensajes
- Usuarios nuevos; La cantidad de personas que se han unido a tus usuarios en cierto tiempo.

Métricas de los mensajes

- Mensajes iniciadores de conversación: Número de veces que el chatbot inicia la conversación con los usuarios.
- Mensajes del chatbot: Número de mensajes que emite el chatbot por conversación.
- Mensajes entrantes: Número de mensajes que han entrado al sistema
- Mensajes fallados: Número de mensajes que el chatbot no ha podido interpretar
- Conversaciones totales: Número de conversaciones iniciadas y completadas satisfactoriamente.
- Conversaciones nuevas: Número de conversaciones nuevas iniciadas por el usuario.

Métricas del chatbot:

- Ratio de retención: el porcentaje de usuarios que regresan a usar el servicio en un marco de tiempo.
- Ratio de cumplimiento de objetivos: El porcentaje de enganchamiento satisfactorio logrado.
- Tiempo/mensajes/intentos para completar el objetivo por parte del chatbot.
- Ratio de retornos: Porcentaje de veces que el chatbot ha fallado
- Satisfacción del usuario: Si el servicio satisface la necesidad del usuario.
- Viralización: Que tanto se comparte y se viraliza el chatbot

Existen diversas métricas y formas de evaluar los chatbots desde el punto de vista del desarrollador o empresa dependiendo de la necesidad y la tarea. Sin embargo, es importante la percepción de los usuarios. Esto se obtiene a través de la usabilidad percibida del usuario. La usabilidad es una variable importante actualmente en este

mar de sistemas debido a que una buena usabilidad puede significar la fidelidad del usuario.

1.3.7. Evaluaciones de usabilidad.

La usabilidad se refiere a la satisfacción total del usuario. A qué tanto una herramienta puede ser utilizada para realizar tareas específicas.[55,56]. Entre las herramientas de medición más populares, se encuentra los cuestionarios: CSUQ, SUS (EUS en español), UMUX y UMUX Lite [57]. La mayor diferencia entre ellos es la cantidad de preguntas y en algunos casos las sub-escalas adicionales que posee cada cuestionario. Por ejemplo, el cuestionario CSUQ evalúa la usabilidad y también da detalles de la usabilidad del sistema en las sub-escalas: utilidad (ayuda para cumplir su propósito), calidad de la información(la información que muestra al usuario), calidad de la interfaz (la interfaz con la que interactúa el usuario) [57]

A continuación se muestran los diversos aspectos que evalúa el CSUQ:

- System Usefulness (Utilidad del sistema):
 - Facilidad de uso: Poco esfuerzo
 - Simplicidad del uso: Pocos pasos, pocas partes
 - Rapidez de uso: Velocidad en la ejecución de la tarea
 - Comodidad: No provoca una incomodidad en el usuario.
 - Facilidad de aprendizaje: No cuesta mucho esfuerzo aprender cómo funciona
 - Rapidez de productividad: Que tan rápido puede hacer al usuario productivo en la ejecución de la tarea o búsqueda de su objetivo con el uso del sistema.

- Information Quality (Calidad de la Información):
 - Comunicación frente a problemas: El sistema se comunica frente a errores para poder explicar cómo solucionar estos errores.
 - Facilidad de recuperación: Que tan rápido y fácil es recuperarse de un error hecho cuando se usa el sistema.

- Claridad de Información: Los mensajes mostrados en el sistema se distingue bien.
 - Facilidad de acceso a la Información: Encontrar cierta información usando el sistema.
 - Efectividad de la información: La información proporcionada ayuda a completar el objetivo
 - Organización de la información: La organización de la información del sistema en la pantalla es clara.
- Interface Quality (Calidad de la Interfaz)
 - El interfaz es placentero: No brinda incomodidad al usuario.
 - Da gusto el uso de la interfaz: El interfaz atrae a su uso.
 - Funcionalidad de la Interfaz: El sistema tiene todas las funciones y expectativas que se espera.
 - Satisfacción del sistema
 - Satisfacción del sistema: Qué tan satisfecho uno después de usar el sistema.

Los modos de evaluar los servicios de bots conversacional han ido cambiando dependiendo del punto de vista de evaluación. Y con el tiempo, ya que se tiene confianza en la calidad de los servicios de IA para el procesamiento de lenguaje natural, la experiencia y retención del cliente/usuario han pasado a primer plano.

1.3.8. Chatbots en diversos ámbitos.

Actualmente, diferentes empresas cuentan con chatbots con diferentes propósitos. Existen chatbots para ayudar a usar dispositivos, hacer llamadas, organizar agendas, registrar contactos, etc. Entre los más conocidos se encuentran:

- Siri (apple.com/es/siri/), para celulares, tablets y PCs de la marca Apple.
- Cortana (microsoft.com/es-es/windows/cortana), incluida en las versiones de Windows de la empresa Microsoft.

- Bixby (samsung.com/es/apps/bixby/), añadida últimamente a los móviles de Samsung.
- El asistente de google (assistant.google.com/), que se encuentra en los dispositivos que funcionan con el sistema Android.
- Alexa (developer.amazon.com/es-ES/alexa), que da soporte dentro de los dispositivos de la empresa Amazon.

También hay chatbots alojados en servicios de mensajería social como Facebook, Whatsapp, en páginas web o en aplicativos de las cuentas de diferentes negocios.

Por ejemplo:

- Para ayudar a agendar vuelos: como es el caso de BB (bb.klm.com/en) de la aerolínea KLM.
- Para saber el número de cuenta, o saldo de cuenta bancaria: existe Arturito de la empresa BCP (m.me/ArturitoBCP).
- Para enseñar diferentes idiomas: está Duolingo que permite ejercitarse con Chatbots (www.duolingo.com/).
- Para brindar información sobre productos: como Kian de la empresa Kia que explica características de diferentes autos (kia.com/us/en/chatbot)

Del mismo modo, en el campo de la salud, recientemente se han comenzado a estudiar el uso de chatbots.

1.3.9. Customer Relationship Management (CRM)

Los CRM son herramientas utilizadas actualmente para la mejora del manejo, enfoque y uso de las relaciones con los clientes. [58]. Se utilizan las tecnologías de la información para optimizar los procesos de la empresa (ej. SAP) [59] o para manejar la comunicación con los clientes (ej. Salesforce) [60].

Por este motivo, unas de las más recientes herramientas que se han añadido a diversos CRM, son los chatbots en servicios de mensajería social [61]. Estos apoyan en la parte más repetitiva del seguimiento. Un ejemplo es “Zoho”, un CRM que tiene como objetivo automatizar las interacciones de compra y venta con los clientes. [62].

Así un sistema, apoyado de servicios automáticos con diversos canales, puede capturar datos de los clientes de forma sistemática a través de conversaciones, o en otras palabras de una forma más natural para el cliente, sin dejar de lado la interacción directa con una persona cuando sea necesario.

1.4. Justificación de la Investigación

El COVID-19 es una enfermedad letal y altamente contagiosa que ha provocado una pandemia y un desgaste adicional en la principal línea de defensa de salud nacional que son los profesionales sanitarios que requieren protección.

En el Perú como parte de la búsqueda de soluciones que pudieran aportar a la protección del personal de salud, a mediados del 2020 se aprueba un ensayo clínico para evaluar el uso de HCQ como profilaxis frente al COVID-19 en estos profesionales. Este estudio requería un seguimiento y monitoreo intenso pero remoto (dadas las condiciones de pandemia) de todos los reclutados a fin de asegurar su participación, recoger información sobre la toma del medicamento, eventos adversos y síntomas que pudieran presentar, manteniendo las condiciones de distanciamiento social. Se requería innovación dentro del ensayo para ayudar en este aspecto.

Actualmente se han desarrollado tecnologías de IA para mantener conversaciones y hacer seguimiento de individuos usando CRM y chatbots para obtener data otorgada por varios clientes de forma automática y simultánea a través de conversaciones. Si bien existen investigaciones que evalúan el uso de algunos chatbots en salud, aún son pocas y mucho menos, de CRM.

Por ello se planteó el uso y evaluación de la usabilidad del chatbot y el CRM dentro del ensayo clínico de profilaxis con HCQ.

2. Objetivos

2.1. Objetivo general

- Evaluar la usabilidad de un chatbot para el seguimiento de los participantes del proyecto “Hidroxiclороquina para prevenir infección por SARS-CoV-2 en personal de salud: Ensayo clínico controlado de fase 3, aleatorizado, de etiqueta abierta” con número de SIDISI 202087 y registro en Clinical Trials NCT04414241 [10].

2.2. Objetivos secundarios

- Describir métricas relacionadas al uso del chatbot
- Evaluar la usabilidad percibida por los participantes en el ensayo clínico y compararla entre grupos diferenciados por edad, sexo, brazo de intervención y tipo de personal de salud.
- Evaluar la usabilidad percibida y la utilidad para el seguimiento en el ensayo por el equipo a cargo de la operación del sistema CRM que interactúan con el sistema, reciben la data y cuando es necesario interactúan con el participante

3. Metodología

3.1. Diseño del Estudio

Este es un estudio descriptivo de usabilidad de un chatbot en un CRM, el cual ha sido diseñado e implementado específicamente para la tarea de seguimiento del ensayo “Hidroxicloroquina para prevenir infección por SARS-CoV-2 en personal de salud: Ensayo clínico controlado de fase 3, aleatorizado, de etiqueta abierta” con número de SIDISI 202087 y registro en Clinical Trials NCT04414241 [10]. Se planeó seguimiento de los participantes por un periodo de aproximadamente 8 semanas o antes si el participante se retiraba del ensayo clínico.

La evaluación se realizó desde dos diferentes tipos de usuarios:

- Los participantes en el ensayo
- El equipo a cargo de la operación del sistema para el monitoreo.

El Chatbot sirvió para monitorear tres aspectos:

- (1) Adherencia a la intervención (toma de HCQ en el grupo de intervención)
- (2) Posibles eventos adversos en ambos brazos del estudio.
- (3) Aparición de síntomas compatibles con COVID-19 en ambos brazos del estudio.

3.2. Población

Esta evaluación se realizó con participantes del ensayo y con miembros del equipo a cargo de la operación del sistema del ensayo clínico.

El ensayo clínico propuso un tamaño de muestra de 320 personas (160 para cada brazo) que trabajen como personal de salud del Hospital Cayetano Heredia, Hospital Nacional Hipólito Unanue (HHU), Hospital Nacional Arzobispo Loayza o el Centro Médico Naval "CMST". Sin embargo, solo lograron reclutar 68 participantes en total en el ensayo clínico. Para nuestra evaluación se invitó a los 68 participantes del ensayo quienes usaron el chatbot como parte de su seguimiento.

3.3. Comunicación con el Chatbot

3.3.1. Funcionamiento diario del sistema chatbot

Con el fin de capturar los datos de la adherencia, los síntomas y eventos adversos, el chatbot llevaba a cabo interacciones y a través de mensajes seguía un flujo de dialogo programado.

- Una interacción es un intento de comunicación entre el chatbot y el usuario que se inicia con un mensaje
- Un mensaje es una porción de texto que se muestra como un globo en el chat que puede ser originado por el chatbot o por el participante.

Diariamente, se seguía este cronograma.

1. A las 08:30, el sistema enviaba un mensaje masivo (a todos los participantes) para iniciar la interacción.
 - En caso de que el participante responda completamente al chatbot (hasta terminar toda la interacción), esta interacción se considera “satisfactoria sin avisos” ya que no se necesitaron avisos adicionales para que responda.
2. Entre las 10:00 y las 12:00, se enviaban avisos a quienes no hubieran respondido la interacción o no la completaron.
 - En caso de responder completamente después de recibir los avisos, esta interacción se considera: “satisfactoria con avisos”.
3. Luego entre las 15:00 y las 16:00, el equipo deliberaba como proceder con quienes no hubiesen respondido. En algunos casos, se enviaban avisos nuevamente o dependiendo del caso se comunicaban directamente con el participante.
 - Todos los casos que no respondieron hasta antes de las 15:00 se consideraban interacciones “no satisfactorias”.

3.4. Instrumentos usados

3.4.1. Registro de Participantes

Se usó datos del registro de participantes en el ensayo clínico “Hidroxicloroquina para prevenir infección por SARS-CoV-2 en personal de salud”[10], para obtener la información de: brazo de pertenencia, hospital de trabajo, profesión, edad y sexo de los participantes.

3.4.2. Escala de Usabilidad del Sistema

Para evaluar la usabilidad percibida por el usuario se utilizó una adaptación del cuestionario de Escala de Usabilidad del Sistema (EUS) [63], SUS en inglés [64]. Este cuestionario tiene dos versiones una original con preguntas negativas, tales como “*encuentro el sistema innecesariamente complejo*”, y una con preguntas todas positivas, como por ejemplo “*encuentro el sistema, simple*”, donde se ha demostrado que se pueden usar indiferentemente con sus cálculos respectivos [63,65]. Para este estudio decidimos usar la versión de preguntas positivas.

El cuestionario EUS está compuesto de 10 enunciados, al cual el usuario responde que tan de acuerdo se encuentra con el enunciado en una escala de enteros del 1 al 5. En esta escala, 1 significa que se encuentra totalmente en desacuerdo, y 5, totalmente de acuerdo[63].

Se muestra a continuación el cuestionario a usarse:

Escala de Usabilidad del Sistema

En el cuestionario a continuación, se mostrarán 10 enunciados a los cuales debe responder con números del **1** al **5** que significan que está **totalmente en desacuerdo** o **totalmente de acuerdo** con el enunciado.

1. Creo que me gustaría utilizar frecuentemente esta chatbot

2. Encontré el chatbot sencillo

3. Pienso que el chatbot es fácil de usar

4. Pienso que podré utilizar este chatbot sin el apoyo de personal técnico.

5. Encontré que varias de las funciones en el chatbot estaban bien integradas.

6. Pensé que había demasiada consistencia en el chatbot.

7. Me imagino que la mayoría de las personas podrían aprender a usar este chatbot muy rápido.

8. Encontré el chatbot muy intuitivo.

9. Me sentí muy confiado (seguro) al utilizar el chatbot.

10. Pude utilizar el chatbot sin tener que aprender nada nuevo.

Existe dos formas en que se puede interpretar los resultados y que vamos a usar en este estudio.

- Una es considerando si los participantes están de acuerdo con el enunciado, cuando marcan un puntaje de 4 o 5. Para los resultados se presentará el porcentaje de participantes que reportan estar de acuerdo con el enunciado.

- La segunda forma es obteniendo el puntaje EUS [64] para el que se sustrae 1 punto a cada respuesta, se suma y se multiplica por 2.5 como se muestra en la fórmula:

$$EUS = ((E_1 - 1) + (E_2 - 1) + \dots + (E_{10} - 1)) * 2.5$$

Este puntaje EUS va del 0 al 100 pero no debe ser entendido como porcentaje ya que un sistema “promedio” tiene una calificación de 68 [66].

En base a este puntaje EUS se pueden interpretar los resultados según tres diferentes escalas. (1) La escala por adjetivos: Superior/Excelente, Bueno/OK, Pobre/lo Peor [67]. (2) La escala de aceptabilidad [68] y (3) la escala NPS (puntaje promotor neto) [113]. Estas escalas se observan en la tabla 1.

Para esta evaluación, usaremos la escala de aceptabilidad, que es la más usada para aplicativos móviles y considera un nivel de “aceptable” a partir de 70 [67].

Tabla 1. Escala EUS e Interpretaciones (Adaptación de measuringu.com)

| Puntaje EUS | Adjetivos | Aceptabilidad | NPS |
|-------------|-------------------------|---------------|-----------|
| 84.0-100 | Superior – Excelente | Aceptable | Promotor |
| 80.7-84.0 | | | |
| 78.8-80.7 | | | |
| 77.1-78.8 | | | Pasivo |
| 74.0 – 77.1 | | | |
| 72.5 – 74.0 | | | |
| 70.0– 72.5 | Ok-Bueno | Marginal | Detractor |
| 64.9 –70.0 | | | |
| 62.6 – 64.9 | | | |
| 51.6 – 62.6 | Lo Peor - Pobre | No Aceptable | |
| 25 – 51.6 | | | |
| 0-25 | | | |

El instrumento de usabilidad (EUS) se aplicó a los usuarios del sistema (participantes del estudio), y también se aplicó al equipo de investigación a cargo de la operación del sistema.

3.4.3. Cuestionario de utilidad del Sistema para el equipo de investigación

Para evaluar la utilidad del sistema por parte del equipo a cargo de la operación no se utilizó un cuestionario estándar de utilidad como el incluido en las sub-escalas del CSUQ[57] o del TAM (Technology Acceptance Model)[70] ya que están más enfocados en herramientas para la producción. Se plantearon preguntas para medir la utilidad de la herramienta de seguimiento en el ensayo clínico y además para recoger posibles mejoras. Este cuestionario incluyó 20 preguntas: 7 preguntas para marcar del 1 al 5 y 13 preguntas abiertas.

Cuestionario de utilidad del sistema

A continuación, se le presentará 20 preguntas relacionadas a la utilidad del sistema.

1. El sistema me permite atender varios participantes a la vez de forma fácil. Desde su punto de vista marque de 1 al 5, donde 1 indica que se encuentra en total desacuerdo y 5, en total acuerdo

Podría explicarnos ¿Cómo le ayudo o no le ayudó a manejar varios participantes a la vez?

Podría explicarnos ¿Cómo le ayudaría a manejar más participantes con menos esfuerzo?

2.El sistema me ahorra todas las tareas repetitivas que podría conllevar el seguimiento de los participantes. Desde su punto de vista marque de 1 al 5, donde 1 indica que se encuentra en total desacuerdo y 5, en total acuerdo

Podría explicarnos ¿Qué tareas repetitivas, le ayudó a solucionar el sistema?

Podría explicarnos ¿Qué tareas repetitivas desearía que le ayude a solucionar el sistema?

3. Siento que los participantes se mostraron dispuestos a usar el sistema. Desde su punto de vista marque de 1 al 5, donde 1 indica que se encuentra en total desacuerdo y 5, en total acuerdo

Podría explicarnos ¿Qué percibió que atrajo a los participantes a usar el sistema?

¿Qué otra funcionalidad, cree usted que podría ayudar a mantener la disponibilidad de usar el sistema en los participantes?

4.Siento que los participantes se aburririeron de usar el sistema. Desde su punto de vista marque de 1 al 5, donde 1 indica que se encuentra en total desacuerdo y 5, en total acuerdo

Podría explicarnos ¿Por qué cree que el sistema aburre o no a los participantes?

Podría explicarnos ¿Cómo podríamos mantener el interés del participante?

5. El sistema permite que pueda monitorear los síntomas de forma segura frente al Covid-19. Desde su punto de vista marque de 1 al 5, donde 1 indica que se encuentra en total desacuerdo y 5, en total acuerdo.

Podría explicarnos ¿Por qué cree que el sistema le permite o no monitorear de forma segura?

6. Siento que el sistema fue seguro para manejar los datos de los participantes. Desde su punto de vista marque de 1 al 5, donde 1 indica que se encuentra en total desacuerdo y 5, en total acuerdo.

Podría explicarnos ¿Por qué cree o no que el sistema es seguro con los datos personales de los participantes?

¿Cómo cree usted que el sistema podría ser más seguro con los datos de los participantes?

7. Siento que el sistema fue útil para el objetivo del ensayo de Hidroxicloroquina. Desde su punto de vista marque de 1 al 5, donde 1 indica que se encuentra en total desacuerdo y 5, en total acuerdo.

Podría explicarnos ¿Por qué cree o no que el sistema fue útil para el ensayo de Hidroxicloroquina?

¿Cómo cree usted que podría haber sido más útil?

3.5. Procedimientos

3.5.1 Diseño del Sistema

El sistema se diseñó siguiendo un flujo de diálogo basado en los requerimientos del ensayo como se muestra en el Anexo I. Este flujo de diálogo se adecuó a las necesidades del estudio. En el flujo se incluyen:

- Para ambos grupos
 - Preguntas sobre síntomas asociados a COVID-19
 - Preguntas sobre posibles efectos adversos de la HCQ
- Para el grupo de intervención
 - Preguntas sobre adherencia de la HCQ.

3.5.2 Piloto del Sistema

Se realizaron simulaciones de las situaciones que podrían encontrar los usuarios.

3.5.3 Construcción en el sistema de “Customer Relationship Management”

Con el funcionamiento del sistema probado y validado, se desarrolló la estructura del sistema considerando un sistema CRM que permita la comunicación en los canales que sean de fácil acceso para los participantes del ensayo y permita el monitoreo remoto.

3.5.4 Entrenamiento para su uso

El sistema, si bien funcionaba de modo semiautomático, necesita de atención directa con el equipo de monitoreo en caso suceda alguna comunicación crítica con el participante. Por este motivo, se vio la necesidad de capacitar al equipo a cargo de la operación del sistema para que sepan cómo manejar la herramienta para la comunicación con el participante del ensayo.

3.5.5. Implementación

La implementación se inició dentro del ensayo clínico. Se adjunta la figura 6 que muestra cómo se incorporó el sistema en el ensayo clínico.

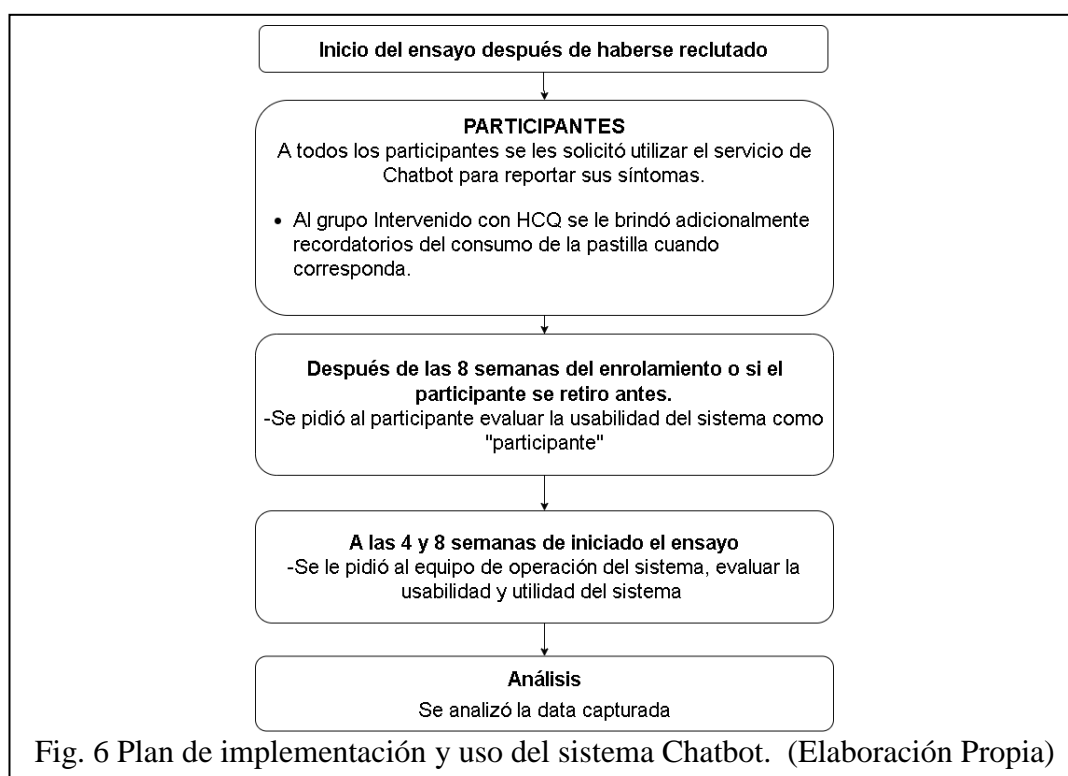


Fig. 6 Plan de implementación y uso del sistema Chatbot. (Elaboración Propia)

En horas de la mañana, los participantes recibieron un saludo para iniciar el monitoreo. El grupo intervenido recibió como mensaje adicional, el recordatorio del uso de la Hidroxicloroquina como profilaxis la fecha que corresponda. Pero a todos se les preguntó sobre los síntomas que puedan sentir diariamente. Se realizaron cambios en la interacción en base a recomendaciones por los participantes o expertos.

3.5.6. Evaluación

Métricas del chatbot

Se colectó diariamente la información sobre las métricas del chatbot por participante, hasta el final de seguimiento del último participante. Se evaluó el tiempo por interacción satisfactoria a través de tres métricas:

- Tiempo de espera (TE): Se considera desde que el chatbot se comunica hasta que recibe un mensaje de respuesta del usuario.
- Tiempo de interacción (TI): Empieza cuando se recibe el mensaje del usuario hasta que se termina toda la interacción con el sistema.
- Tiempo total de uso de sistema por participante (TT): Comprende la suma de tiempo de espera y tiempo de interacción.

Usabilidad percibida por los participantes

Luego de las 8 semanas de seguimiento; o antes, si su participación culminó según razones establecidas en el estudio, se invitó al participante a través del chatbot a que evalúe la usabilidad del sistema de forma electrónica.

A los participantes se les mostró el siguiente mensaje por chat

Buenos días, queremos invitarlo a participar en una investigación para evaluar este sistema de chatbot. El estudio se denomina “Evaluación del uso de un Chatbot para el seguimiento de un ensayo clínico de profilaxis frente al COVID-19 en personal de salud”, y su participación consiste en responder a 10 preguntas, que le tomarán solo unos minutos en responder. No registraremos datos personales

Si desea más información puede encontrarla en el siguiente enlace: *<enlace a hoja informativa>*

Si desea participar por favor responda SI o NO
(Si responde NO, se le manda un mensaje: “Muchas gracias por su tiempo”)
(Si responde SI, se le envía la explicación del cuestionario y luego aparecerá pregunta por pregunta en el chat)

Luego de esto, se le explicó al usuario como contestar y los 10 enunciados fueron apareciendo hasta que acabe el cuestionario.

En el cuestionario a continuación, se mostrarán 10 enunciados a los cuales debe responder con números del 1 al 5 donde 1 significa que está **totalmente en desacuerdo** y 5 que está **totalmente de acuerdo** con el enunciado.

Creo que me gustaría utilizar frecuentemente esta chatbot

Encontré el chatbot sencillo

Pienso que el chatbot es fácil de usar

Pienso que podré utilizar este chatbot sin el apoyo de personal técnico.

Encontré que varias de las funciones en el chatbot estaban bien integradas.

Pensé que había demasiada consistencia en el chatbot

Me imagino que la mayoría de las personas podrían aprender a usar este chatbot muy rápido.

Encontré el chatbot muy intuitivo.

Me sentí muy confiado (seguro) al utilizar el chatbot

Pude utilizar el chatbot sin tener que aprender nada nuevo

Muchas gracias por participar del estudio. ¡Hasta pronto!

La data se ingresó a la base de datos del sistema y luego fue analizada usando Epiinfo y R.

Usabilidad y utilidad percibida por el equipo a cargo de la operación del sistema

La evaluación del equipo a cargo de la operación del sistema se realizó en dos momentos. En un primer momento, a las 4 semanas, de haber iniciado el ensayo, se les envió una invitación por correo junto con la hoja informativa del estudio y un enlace al formulario electrónico. En un segundo momento, a las 8 semanas, se les envió un segundo correo a evaluar de nuevo la usabilidad y utilidad del sistema como se muestra a continuación.

Correo electrónico:

Estimado <nombre del operador>

Buenos días, queremos invitarlo a participar en una investigación para evaluar este sistema de chatbot. Su participación es muy importante por el rol que ocupa en el Ensayo Clínico. Este estudio se denomina “Evaluación del uso de un Chatbot para el seguimiento de un ensayo clínico de profilaxis frente al COVID-19 en personal de salud”, y su participación consiste en responder un cuestionario de 10 preguntas con una escala de 1 a 5 y 20 preguntas abiertas ahora y al final del ensayo clínico. Estimamos que cada vez el cuestionario le tomará unos 15 a 20 minutos. No registraremos datos personales. La información del estudio se puede encontrar en el archivo “*hoja informativa*” que se encuentra adjunto.

Si Ud. desea participar haga clic en este *enlace* y proceda con el cuestionario.

Muchas gracias por su colaboración.
Anthony Bacilio Ruiz

La data fue ingresada a una base de datos para su análisis. Para el segundo cuestionario al cabo de 8 semanas se envió un mensaje similar a todos los que accedieron a participar en el primer cuestionario con el enlace al instrumento.

3.6. Análisis de la data

Para el análisis de la data se utilizó el paquete Epi info (Versión 7.2.4.0) [71] y Rstudio(V 1.4.1717)[72]. Se calcularon variables descriptivas de los participantes y las métricas de uso y se usaron frecuencias, porcentajes, medias y/o desviaciones.

De acuerdo a las variables, se evaluaron si cumplían los supuestos de normalidad, homogeneidad de varianza, linealidad y distribución similar para definir los procedimientos estadísticos a usarse.

Para la usabilidad percibida por los usuarios se reporta el “porcentaje de acuerdo” para cada enunciado de la encuesta. Este es el porcentaje de participantes que marcaron 4 o 5 en el enunciado. Adicionalmente se calculó y reportó el promedio total del “porcentaje de acuerdo”.

El puntaje EUS (escala de usabilidad del sistema) de cada participante se calculó según lo descrito en la pagina 27. Luego se halló el promedio del puntaje EUS o “usabilidad promedio”.

Se evaluó la correlación entre el puntaje EUS y la edad a través de la prueba de correlación de Spearman. Para evaluar la relación entre el puntaje EUS y las otras características sociodemográficas se uso la prueba de Kruskall Wallis .

Los datos de la usabilidad y utilidad por parte del equipo de operación del sistema, están presentados de manera descriptiva de los dos momentos de medición (4 semanas vs. 8 semanas). Por último, se presentan los comentarios recibidos y las mejoras que lograron implementarse.

3.7. Consideraciones éticas

El protocolo y todos los instrumentos del estudio fueron evaluados y aprobados por el comité de ética de la Universidad Peruana Cayetano Heredia. Los participantes del ensayo fueron invitados a participar en la evaluación de la usabilidad, se le dio acceso a la hoja informativa y si marcaban que querían participar, podían completar el cuestionario. En el caso del equipo de operación, se les invitó a ser parte del estudio por correo, adjuntándoles la hoja informativa. Se incluyó, en el correo el enlace al cuestionario.

4. Resultados

4.1 Características demográficas de los participantes del ensayo

El ensayo clínico enroló a 68 participantes en total. Todos ellos fueron invitados a la evaluación de usabilidad del sistema, pero solo 42 (61.8%) aceptaron participar. De estos, 2 no llenaron la encuesta. De modo que el análisis de usabilidad se realizó con 40 de los 68 participantes del ensayo (58.8%). Las características demográficas de los participantes del ensayo se presentan en la Tabla 2. En el grupo que aceptó participar el 57% fueron mujeres. La edad promedio fue de 39, (rango 23-56 años). La mayoría fueron médicos (64.3%), y la sede con mayor número de participantes fue el Hospital Naval (54.8%).

Tabla 2: Características demográficas de los participantes (N=40)

| Características sociodemográficas | | |
|--|--------------|--------------|
| | media | rango |
| Edad en años | 39 | (23-56) |
| Sexo | N | % |
| Femenino | 22 | 55 |
| Profesiones | | |
| Médico | 25 | 62.5 |
| Enfermero | 8 | 20 |
| Técnico | 6 | 15 |
| Auxiliar | 1 | 2.5 |
| Hospitales | | |
| H. Cayetano | 17 | 42.5 |
| H. Naval | 21 | 52.5 |
| H. N. A. Loayza | 2 | 5 |
| Brazo del ensayo | | |
| Control | 21 | 52.5 |
| Intervención | 19 | 47.5 |

4.2. Métricas de usuarios y del chatbot

El sistema del chatbot, estuvo en funcionamiento activo 156 días, desde el día 1 del primer reclutado (25 de junio) hasta el final de seguimiento del último participante (27 de noviembre). Dado que fueron 68 participantes y el seguimiento era por 8 semanas máximo. Se esperaba un total de 3808 interacciones iniciadas por el chatbot; sin embargo, varios reclutados terminaron su participación antes de las 8 semanas de modo que se registraron solo 3526 interacciones de las cuales el 86.6% fueron consideradas satisfactorias. (Ver tabla 3).

Tabla 3. Métricas de las interacciones del chatbot

| Métricas sobre las interacciones | Valor | Acumulado | % |
|--|--------------|------------------|----------|
| Interacciones satisfactorias sin avisos | 762 | 762 | 21.6% |
| Interacciones satisfactorias con avisos | 2290 | 3052 | 65.0% |
| Interacciones insatisfactorias | 474 | 3526 | 13.4% |
| Total | 3526 | 3526 | 100% |

Como resultado de todas las interacciones, se intercambiaron un total de 26,903 mensajes. El chatbot creó 11088 mensajes; los monitores, 7801; y los usuarios. 8014.

Las métricas con respecto a los tiempos de interacciones satisfactorias, se muestran en la tabla 4. La mediana del tiempo total (TT) fue de 33.0 y 61.8 minutos para los grupos de intervención y control. La mediana del tiempo de interacción (TI) fue de 14 y 28 segundos respectivamente. Responder al sistema no toma mucho tiempo. El tiempo más largo fue el tiempo de espera de respuesta (TE). Muy probablemente los participantes luego de recibir el mensaje, esperaban el momento adecuado para

comenzar a responder. Unos pocos, tomaron hasta más de 5 horas en la interacción, ya que no se dio de manera continua.

Tabla 4. Métricas del tiempo usado por participante.

| Métricas sobre el tiempo usado en las interacciones satisfactorias | Controles | Intervención | Total |
|---|------------------|---------------------|--------------|
| Tiempo de Espera | | | |
| Tiempo mínimo | 5 s | 6 s | 5 s |
| Tiempo máximo | 6.4 horas | 6.4 horas | 6.4 horas |
| Mediana | 29.6 min | 40.6 min | 32.6 min |
| Promedio | 1.3 horas | 1.5 horas | 1.4 horas |
| Tiempo de Interacción | | | |
| Tiempo mínimo | 0 | 0 | 0 |
| Tiempo máximo | 5.9 horas | 6.4 horas | 6.4 horas |
| Mediana | 14 s | 28 s | 19 s |
| Promedio | 4.7 min | 17.8 min | 10.6 min |
| Tiempo Total | | | |
| Tiempo mínimo | 5 s | 14 s | 5 s |
| Tiempo máximo | 6.4 horas | 6.4 horas | 6.4 horas |
| Mediana | 33.0 min | 61.8 min | 43.6 min |
| Promedio | 1.3 horas | 1.8 horas | 1.6 horas |

Podríamos estimar que, si la tarea de contactar a los participantes y completar la información sobre adherencia y síntomas hubiera sido realizada por un monitor, este podría haber tenido que invertir alrededor de 43.6 minutos (que es la mediana del TT por interacción para ambos grupos) por cada una de las 3052 interacciones satisfactorias del estudio (86.56% de las 3526). Entonces el total resultaría en 2219.82 horas (aproximadamente 92 días de trabajo) que se han ahorrado gracias al sistema. El sistema permite ahorrar trabajo a los monitores dado que realiza la tarea de enviar mensajes de forma automática.

4.3. Evaluación de usabilidad por parte de los participantes

El promedio del puntaje obtenido a través del cuestionario EUS (escala de usabilidad del sistema) fue de 78.25 (D.E. 24.43), lo que en la escala de usabilidad se encuentra muy por encima del límite de aceptabilidad, siendo el chatbot considerado como la categoría más alta: “Aceptable” (ver tabla 1 en metodología).

En la tabla 5 se presentan los “porcentajes de acuerdo” con los enunciados sobre usabilidad. En todos los enunciados la mayoría de los participantes estuvieron de acuerdo, dando un porcentaje de acuerdo promedio de 77.5%. El 72.5% creen que podrían usar frecuentemente el chatbot, el 85% reporta no necesitar apoyo técnico, 80% no necesitaron aprender algo nuevo y el 92.5% está de acuerdo que nuevos usuarios pueden aprender rápido.

Sobre el chatbot la gran mayoría opinó que es sencillo (90%), fácil de usar (90%), y seguro (82.5%). Sin embargo, solo un poco más del 52% consideró que era consistente, el 62.5% que era bien integrado y el 67.5% que era intuitivo.

Tabla 5: Porcentaje de participantes que están de acuerdo con cada enunciado (n=40)

| No | Ítem | % de acuerdo |
|----|---|--------------|
| 1 | Creo que me gustaría utilizar frecuentemente esta chatbot | 72.5 |
| 2 | Encontré el chatbot sencillo | 90.0 |
| 3 | Pienso que el chatbot es fácil de usar | 90.0 |
| 4 | Pienso que podré utilizar este chatbot sin el apoyo de personal técnico | 85.0 |
| 5 | Encontré que varias de las funciones en el chatbot estaban bien integradas. | 67.5 |
| 6 | Pensé que había demasiada consistencia en el chatbot. | 52.5 |
| 7 | Me imagino que la mayoría de las personas podrían aprender a usar este chatbot muy rápido | 92.5 |
| 8 | Encontré el chatbot muy intuitivo. | 62.5 |
| 9 | Me sentí muy confiado (seguro) al utilizar el chatbot. | 82.5 |
| 10 | Pude utilizar el chatbot sin tener que aprender nada nuevo. | 80.0 |
| | Promedio de porcentaje de acuerdo | 77.5 |

Luego se evaluó si existían diferencias en el puntaje EUS según las características sociodemográficas (ver Tabla 6). Al usar la prueba de Spearman, no se encontró correlación con la edad. Por la prueba de Kruskal-Wallis, tampoco se encontró que existan diferencias según alguna de las otras características sociodemográficas.

Tabla 6: Evaluación del Puntaje EUS como variable dependiente (n=40)

| Característica | Media (EUS) | RIQ (EUS) | p-value |
|--------------------------------|--------------------|------------------|----------------|
| Edad (años) | 78.25 | 22.5 | 0.814* |
| Sexo | | | 0.074** |
| Mujer | 73.52 | 23.1 | |
| Hombre | 84.03 | 23.8 | |
| Brazo del Ensayo | | | 0.115** |
| Intervención | 76.58 | 17.5 | |
| Control | 79.76 | 25.0 | |
| Profesiones (Agrupado): | | | 0.192** |
| Médico | 80.4 | 25.0 | |
| No Médico | 74.67 | 17.5 | |
| Profesiones: | | | 0.484** |
| Médico | 80.4 | 25.0 | |
| Enfermero | 74.38 | 19.4 | |
| Técnico | 74.17 | 8.8 | |
| Auxiliar | 80 | - | |
| Hospitales | | | 0.320** |
| H. Cayetano | 84.52 | 20.0 | |
| H. Naval | 74.56 | 30.0 | |
| H. N. A. Loayza | 43.75 | - | |

*Prueba de Spearman,

**Prueba de Kruskal-Wallis

4.4. Usabilidad y utilidad por parte del equipo a cargo de la operación del sistema (monitores)

4.4.1. Evaluación de Usabilidad

En la evaluación de usabilidad de las semanas 4 y 8 participaron 4 y 3 monitores respectivamente. La mayoría de los “porcentajes de acuerdo” de los enunciados aumentaron entre ambas tomas (Tabla 7). El puntaje total de EUS aumentó con el uso entre las 4 y 8 semanas de evaluación de 65 a 71.7.

Tabla 7: Porcentaje de Monitores que están de acuerdo con los enunciados de usabilidad

| No | Ítem | Primera | Segunda |
|--|---|----------------------------------|----------------------------------|
| | | evaluación a las 4 semanas (n=4) | evaluación a las 8 semanas (n=3) |
| | | % | % |
| 1 | Creo que me gustaría utilizar frecuentemente esta chatbot | 50 | 66.7 |
| 2 | Encontré el chatbot sencillo | 100 | 66.7 |
| 3 | Pienso que el chatbot es fácil de usar | 75 | 100 |
| 4 | Pienso que podré utilizar este chatbot sin el apoyo de personal técnico | 0 | 33.3 |
| 5 | Encontré que varias de las funciones en el chatbot estaban bien integradas. | 75 | 100 |
| 6 | Pensé que había demasiada consistencia en el chatbot. | 50 | 66.7 |
| 7 | Me imagino que la mayoría de las personas podrían aprender a usar este chatbot muy rápido | 100 | 100 |
| 8 | Encontré el chatbot muy intuitivo. | 75 | 100 |
| 9 | Me sentí muy confiado (seguro) al utilizar el chatbot. | 50 | 66.7 |
| 10 | Pude utilizar el chatbot sin tener que aprender nada nuevo. | 50 | 66.7 |
| Promedio de porcentaje en acuerdo | | 62.5 | 76.7 |

4.4.2. Evaluación de utilidad

De este cuestionario se obtuvo la percepción de cuánto apoyó el chatbot al equipo de monitores del ensayo clínico y qué aspectos se deben mejorar.

El porcentaje en acuerdo con los enunciados aumentó de un 67.9% a un 71.4%. Por otro lado, ningún monitor reportó sentir que el chatbot aburriría a los participantes y más bien todos coincidieron que el sistema fue útil para el ensayo clínico (Tabla 8).

Tabla 8 Resultados de evaluación de utilidad (Porcentaje de acuerdo con los enunciados)

| No | Ítem | Primera evaluación a las 4 semanas (n=4) % | Segunda evaluación a las 8 semanas (n=3) % |
|--|--|---|---|
| 1 | El sistema me permite atender varios participantes a la vez de forma fácil. | 100 | 66.7 |
| 2 | El sistema me ahorra todas las tareas repetitivas que podría conllevar el seguimiento de los participantes | 75 | 66.7 |
| 3 | Siento que los participantes se mostraron dispuestos a usar el sistema. | 75 | 100 |
| 4 | Siento que los participantes se aburrieron de usar el sistema. | 0 | 0 |
| 5 | El sistema permite que pueda monitorear los síntomas de forma segura frente al Covid-19. | 75 | 100 |
| 6 | Siento que el sistema fue seguro para manejar los datos de los participantes. | 50 | 66.7 |
| 7 | Siento que el sistema fue útil para el objetivo del ensayo de Hidroxicloroquina. | 100 | 100 |
| Promedio de porcentaje en acuerdo | | 67.9 | 71.4 |

De las preguntas abiertas se recogieron comentarios y recomendaciones. Por ejemplo, los miembros del equipo comentaron que el sistema permitió tener reportes diarios de la mayoría de los pacientes y enfocarse en el seguimiento de los que tenían problemas o síntomas. También sus recomendaciones ayudaron a mejorar el sistema. Uno de los problemas vistos fue con el uso del iPad que, al tener una resolución considerada como dispositivo móvil por el navegador web, disponía los elementos de la ventana de una manera diferente que, con la computadora de escritorio, dificultando algunas tareas. Por otro lado, se solicitó que el sistema maneje una base de datos que sea accesible para los monitores, quitar palabras innecesarias, enviar recordatorios a quienes no responden y hacer más fácil el aprendizaje de nuevas actualizaciones. Todo esto se fue ajustando en el sistema entre la semana 4 y 8 del estudio.

“Si el bot pudiese manejar una especie de base datos recopilando las respuestas de los participantes nos ayudaría en el procesamiento de la información.”

-Monitor 1, 4 semanas

“Nos permitió hacer seguimiento de la toma de pastilla y de los síntomas de un gran número de participantes en simultáneo.”

-Monitor 1, 8 semanas

Acercas de las tareas repetitivas, los monitores resaltaron que el sistema ayudó a indagar a todos los participantes por sus síntomas, y se sugirieron otras funcionalidades como tener envíos de mensajes predeterminados.

“Preguntar por síntomas y por toma de la pastilla todos los días y más de una vez durante el mismo día.”

-Monitor 1, 4 semanas

“Por ejemplo enviar el "no se olvide de asistir a su control" ayuda mucho para solo enviarlo a ciertas personas”

-Monitor 3, 8 semanas

Los monitores percibieron que el sistema atrae a los participantes por su facilidad y simplicidad y por que está basado en WhatsApp que es muy popular.

“Que las respuestas son rápidas, solo con números y a través de WhatsApp que es una aplicación que todos tienen en su celular.”

-Monitor 3, 4 semanas

Los encuestados hicieron notar que el sistema “no aburre” por ser ágil y por no tomar mucho tiempo. También solicitaron que se pudieran mandar algunas imágenes o mensajes motivacionales y sobre el uso de EPP, lo que si se logró implementar.

“(No aburre...) Porque no toma mucho tiempo”

-Monitor 2, 8 semanas

“El uso de imágenes ayuda bastante, sin embargo, sobrecargar con muchas imágenes o quizás una sola imagen con mucha información sobrecargada podría hacer que el participante no le tome interés.”

-Monitor 3, 8 semanas

Respecto a la seguridad de los datos de los participantes, se indicó un punto a mejorar en el sistema. Al registrarse en Whatsapp, los usuarios pueden escoger su “nombre de usuario de Whatsapp” el cual pueden ser conjuntos alfanuméricos, emojis, etc. Luego al interactuar con el chatbot, este “nombre de usuario de Whatsapp” era automáticamente exportado al CRM para identificar cada usuario.

El problema surge cuando algunos participantes escogieron como “nombre de usuario de Whatsapp”, su nombre propio. Si bien el sistema era solo manejado por los monitores del ensayo clínico, se recomendó que se utilicen otro nombre diferente para mantener la confidencialidad.

“El sistema los reconoce por su usuario en Whatsapp que muchas veces puede evidenciar la identidad del participante.”

-Monitor 4, 4 semanas

“Solo los monitores tenemos acceso a la información a través de una cuenta con contraseña personal.”

-Monitor 1, 8 semanas

“Uso de códigos para cada participante”

-Monitor 3, 8 semanas

Por último, sobre la utilidad del chatbot para el ensayo resaltaron que permitió monitorizar de forma efectiva a los participantes en esta situación con poco contacto real dada la pandemia y que fue muy sencillo de usar.

“Nos permite monitorizar de manera efectiva el cumplimiento del tratamiento y los síntomas de todos los participantes.”

-Monitor 1, 4 semanas

Entre la cuarta y la octava semana se implementó el envío de mensajes predeterminados, se añadió el recordatorio de visitas semanales y las demás recomendaciones.

5. Discusión

La usabilidad se refiere a la satisfacción total del usuario. En otras palabras, indica qué tanto una herramienta puede ser utilizada para realizar tareas específicas por percibirse como útil y adecuada [55,56]. En esta investigación, nos planteamos evaluar un chatbot para el apoyo del seguimiento de adherencia y síntomas en participantes en un ensayo clínico. Se lograron más de tres mil quinientas interacciones con el chatbot durante el periodo del estudio y casi el 90% fueron consideradas satisfactorias, es decir el usuario contestó toda la información requerida. Casi la totalidad de los participantes consideraron al chatbot sencillo y fácil de usar, y usando la escala de usabilidad del sistema, el chatbot fue evaluado por los participantes en la categoría más alta de “Aceptable”. El equipo de investigación también evaluó de manera muy positiva la usabilidad y además se documentó que su uso significó un ahorro de más de 2200 horas de trabajo para el equipo del estudio que pudo invertir las en el seguimiento de los pacientes con dificultades.

No hay aún mucha evidencia sobre el uso de herramientas como los chatbots para temas relacionados a salud y existen pocas investigaciones sobre el impacto, calidad o percepción del público de estos chatbots. Más aún, en nuestra búsqueda en la literatura, no hemos encontrado publicaciones sobre el uso de chatbots para apoyo en investigaciones.

Sin embargo, entre los ejemplos de chatbots en el campo de la salud sí se describen diversos usos. Por ejemplo, la fundación Ovarian Cancer Canada utilizó un chatbot en Facebook Messenger para guiar la donación de fondos para la fundación. El

chatbot ofrecía los números de cuenta e información sobre el uso posible de dichas donaciones. [73]

Para el caso de apoyo al diagnóstico, *Your.MD* (en español, tu médico) era un aplicativo que contaba con un chatbot. El usuario digitaba los síntomas y el sistema respondía con diagnósticos, consejos y/o recomendaciones [74]. También existió un servicio de chatbot denominado ADA, la cual atendía casos de problemas de salud mental. A través de una conversación sobre los síntomas, respondía con diagnósticos probables [75]. Por otro lado, existió el chatbot ofrecido en la página de Baby Centre UK de la empresa Johnson & Johnson enfocado en el embarazo y los bebés recién nacidos [76,77]. Este chatbot ayudaba a los padres con recordatorios y consejos frente algunos problemas que podrían suceder en el embarazo y recomendaciones para los recién nacidos como el tipo de alimentación. Lamentablemente ninguno de estos chatbots cuenta con evaluaciones de satisfacción y menos de usabilidad.

Un ejemplo de uso y evaluación de chatbot es la evaluación que se hizo sobre el chatbot Vik [78]. Este se ofrecía a través de aplicativos móviles o web para resolver consultas de pacientes de cáncer de mama. Además, realizaba un cuestionario, donde se respondía con un “*estoy de acuerdo o no estoy de acuerdo*”, una o dos veces a la semana. Se analizaron las conversaciones compiladas en un año y se obtuvo que se presentaba más adherencia a la medicación mientras más uso se le daba al chatbot. También, se encontró que la satisfacción con respecto al uso del chatbot, fue alta (93.9%) y el 88% consideró que Vik les dio soporte y ayuda para seguir el tratamiento efectivamente. Otro ejemplo es Woebot [79], un chatbot que

conversa con los pacientes sobre su estado de ánimo y brinda educación emocional para combatir la ansiedad. A través de un ensayo clínico aleatorizado, se evaluó el efecto del uso de Woebot en comparación con un libro para el control de la ansiedad y depresión. Los resultados mostraron que los participantes del grupo intervenido con el chatbot redujeron significativamente los síntomas de depresión en el tiempo de estudio, medido a través de la herramienta PHQ-9.

Por otro lado, durante la pandemia, Las medidas impuestas por los gobiernos como el distanciamiento social han motivado el desarrollo de soluciones digitales que permitan las operaciones remotas de los servicios de salud, las cuales ahora incluyen chatbots[80–82]. En Latino América se han desarrollado diferentes chatbots. En Uruguay se habilitó un chatbot en la línea 0800-1919 [83–87], Argentina desarrolló a “Boti” en Buenos Aires [88–92], Andy en Neuquén [93] y un chatbot contra la infodemia [94]. Asimismo, “Samibot” apareció en Bolivia [95], CovidChile se desarrolló en Chile[96] y Catalina en Ecuador [97]. Sus funciones fueron principalmente solucionar preguntas frecuentes, pero en algunos casos llamaban ambulancias, hacían telediagnóstico e incluso atendía casos de violencia familiar. Sin embargo, hasta el momento no existe evidencia de evaluaciones a estas soluciones informáticas.

A nivel global, especialmente Europa y Asia han implementado algunos chatbots también durante la pandemia. Unos fueron diseñados como herramientas de diagnóstico para Covid-19 a través de la conversación [98–102]. Otros para cubrir la necesidad de atender preguntas frecuentes sobre el virus, cómo protegerse [100,103–109] u otorgar información combatiendo la desinformación

[100,110,111]. También hubieron chatbots para el apoyo de la atención médica [100,112], pronóstico [113]; vigilancia y censos [102,114] y hasta generación de contenido para mensajes automáticos para redes sociales[115].

En nuestro caso, los participantes del ensayo clínico evaluaron el chatbot dándole un puntaje superior en usabilidad EUS. Hemos encontrado solo algunos ejemplos de uso de EUS para evaluar la aceptabilidad en chatbots sobre salud. KBOT es un chatbot para brindar soporte en el asma infantil y fue evaluado con la escala EUS. Obtuvo un puntaje de 80, cercano al puntaje encontrado en nuestro estudio, representando muy buena aceptabilidad [116]. Otro sistema denominado 3MR_2 fue un chatbot para ayudar a reconstruir pasajes traumáticos y reportó una aceptabilidad con un puntaje EUS entre 73 y 75 [117]. Un par más de chatbots, “Mylo” y “Eliza”, diseñados para ayudar a combatir el estrés en personas adultas no fueron muy bien evaluados por los usuarios usando esta misma escala (Puntaje EUS de 63.56 y 56.97 respectivamente), mostrando la necesidad de cambios y la importancia de medir usabilidad cuando se implementan nuevas estrategias[118].

También se ha publicado un par de apps que promueven adherencia a medicamentos y reporte de síntomas, aspectos que dentro del estudio evaluaba el chatbot. Uno de estos aplicativos fue “HOPE” que se diseñó para apoyar la adherencia a un medicamento para manejar el desorden del consumo de opioides (MAT o Medication-assistent treatment). Como parte de la evaluación obtuvieron el puntaje de aceptabilidad que fue de 86.9 [119]. Otra interesante app fue el BP Journal, una app para controlar la adherencia a medicación contra la hipertensión que también fue evaluado con Puntaje EUS 89.1.

La usabilidad puede ser influenciada por las características del usuario, por ejemplo, por edad o sexo[120,121]. Sin embargo, en nuestro estudio no encontramos una diferencia significativa de la usabilidad ni por edad, sexo, profesión, sede o grupo de intervención, lo que podría ser real o estar relacionado a nuestro limitado tamaño muestral.

Otro aspecto importante de la evaluación de usabilidad con el cuestionario EUS es que nos permite conocer la apreciación sobre las características del sistema a través de sus diferentes ítems. Seis de estos ítems, se refieren a la facilidad con la que se puede realizar las tareas con el sistema: *“encontré el chatbot sencillo”*; *“pienso que es fácil de usar”*; *“es intuitivo”*; *“la mayoría podría aprender a usar este sistema muy rápido”*; *“sin ayuda del personal técnico”* o *“sin aprender nada nuevo”*. En nuestro caso, en estos ítems la gran mayoría considero que era fácil de usar, no necesitan apoyo técnico y no necesitaban aprender nada nuevo. La población a la que este bot ha sido dirigido es una población selecta de profesionales de la salud bastante acostumbrados al uso de celulares. Además, el usar bot en un aplicativo que se ha hecho muy popular en nuestro medio, el WhatsApp, probablemente influyó en su facilidad de uso. Los resultados sobre la facilidad de uso concuerdan con lo obtenido en una evaluación a tres chatbots populares Siri, Alexa y el Google Assistant (En promedio, obtuvieron más de 3.5 de 7 puntos) [122]. Asimismo, con MobileCoach, un chatbot para dialogar sobre enfermedades no comunicables (obtuvo 6.7 de 7 puntos en facilidad de uso)[123] y Elizabeth, un chatbot para dialogar sobre síntomas depresivos (obtuvo 1.99 puntos donde 1 significaba que era fácil y 7, difícil)[124].

En una investigación cualitativa sobre las experiencias de Investigadores usando tecnologías móviles dentro de ensayos clínicos en Estados Unidos, los entrevistados reportaron que uno de los aspectos a destacar al usar tecnologías móviles (no se enfocaron en chatbots), era la posibilidad de la mejora de la eficiencia de los procedimientos de estudio[125]. Ellos se refirieron a la capacidad para monitorear de manera más objetiva el cumplimiento de los procedimientos del estudio y la adherencia a la medicación; el acceso a los datos en tiempo real y a la mejor gestión de los participantes del ensayo. Sin embargo, estas ventajas no hicieron notar el tiempo que se puede ahorrar en seguimiento efectivo, como lo vimos en nuestro estudio.

En cambio, los investigadores compartieron su preocupación por el tiempo necesario para configurar tecnologías y vincular dispositivos a usuarios específicos, el tiempo que necesita el personal del equipo para aprender a usar los sistemas y solucionar problemas técnicos con las apps si funcionan mal. Coincidentemente, los miembros del equipo del ensayo peruano refirieron sobre la necesidad de apoyo técnico con el app. Un aspecto a destacar es que la interfaz del equipo de ensayo clínico en nuestro estudio era diferente a la de los usuarios (no se conectaba a través de WhatsApp, sino que podría hacerlo a través de la web o su celular) y por lo tanto fue necesaria la capacitación en el uso del sistema y la disponibilidad de soporte técnico para ellos.

Si bien el casi el 90% de las interacciones con el chatbot fueron consideradas satisfactorias al ser completadas por los usuarios, es importante destacar que se requirió en al menos en 2/3 de las interacciones un aviso adicional para que el

usuario respondiera. Entonces, en el desarrollo de estos sistemas de chatbot es importante considerar los avisos recordatorios. En un estudio sobre un chatbot RC (Reflection Companion)[126], al momento de reclutar, se enviaba un cuestionario. Luego como seguimiento, se enviaban diariamente mensajes para que el usuario reflexione acerca de su actividad física y en caso de dejar la conversación a medias, se enviaban recordatorios. En este estudio se encontró que los usuarios respondieron en un 96% las preguntas iniciales y en un 90 % las preguntas de seguimiento diarias y se requirió enviar avisos recordatorios a un 39% de las interacciones para que completaran las preguntas.

En nuestro estudio, menos de la quinta parte de las interacciones requirió la intervención adicional por parte del equipo de monitores (llamadas, más avisos o mensajes de WhatsApp, etc), de modo que el sistema fue bastante eficiente reduciendo la carga de trabajo y colectando información a tiempo real.

La usabilidad varía con el tiempo debido a los efectos del aprendizaje, aunque nunca llega a ser perfecto. Una investigación de usabilidad que evaluó semanalmente por 8 semanas como varía el expertise y los niveles de frustración de profesores frente a un software para crear páginas web, halló que el expertise aumenta y la frustración disminuye[127]. Asimismo otra investigación sobre la usabilidad de un sistema de información en salud encontró que los usuarios expertos experimentan menos problemas críticos que usuarios novatos pero sin embargo, aun siguen encontrando problemas serios con los que han convivido por largo tiempo y se deben solucionar[128].

En nuestro estudio se hicieron evaluaciones en dos tiempos, 4 y 8 semanas. Observamos un incremento en el puntaje de usabilidad que podría deberse al proceso de aprendizaje, es decir luego de varias semanas estaban más adecuados al sistema y eso podría “mejorar” su puntaje de usabilidad.

Una de las limitaciones mas importantes de nuestro estudio es el reducido tamaño de muestra, a causa de la escasez de enrolados en el ensayo clínico original (68). Esta situación limita el poder de nuestro estudio. Por ejemplo, para hacer análisis comparando sexos y puntajes de EUS y asumiendo distribución normal, el estudio no tendría suficiente poder para detectar como estadísticamente significativas diferencias menores a 22 puntos entre los 18 hombres y las 22 mujeres que participaron. La diferencia observada fue de 10.51 puntos.

A pesar de las limitaciones, todos los participantes del ensayo clínico recibieron las interacciones y más del 50% completaron los cuestionarios de evaluación finales. Este chatbot se desarrolló específicamente para apoyar en el ensayo en la tarea de evaluar síntomas y adherencia al tratamiento. Sin embargo, la misma lógica de programación y presentación de la información podría adecuarse y usarse para otros ensayos en los que se requiera seguimiento y data a tiempo real. Finalmente, consideramos que nuestro estudio aporta a la todavía poca, pero ahora creciente evidencia de la usabilidad de los chatbots en salud y en investigación para la salud; y sobre el rol de las tecnologías de la información en mejorar la eficiencia y reducir la carga de trabajo.

6. Conclusiones

Se evaluó la usabilidad de un chatbot para el seguimiento de participantes de un ensayo clínico y se obtuvo un puntaje de usabilidad alto lo que indica que el sistema fue muy bien evaluado. No se encontraron diferencias según las características sociodemográficas de los participantes de esta evaluación.

La gran mayoría de las interacciones fueron satisfactorias, y permitieron que se ahorren más de dos mil horas de trabajo. Solo 13.4% necesitaron una intervención mayor del equipo de monitores debido a que los participantes del ensayo no contestaron los chats del chatbot. Esto permitió ahorrar tiempo e hizo más efectivo el estudio, por lo que los monitores calificaron el sistema también con un puntaje de usabilidad alto y consideraron que era útil para el seguimiento de los participantes.

Este tipo de aplicaciones podrían adecuarse y usarse más frecuentemente en investigación y también en otras aplicaciones de salud que requieran seguimiento de síntomas y adherencia al tratamiento, como enfermedades crónicas y otras.

7. Referencias bibliográficas

1. Palacios Cruz M, Santos E, Velázquez Cervantes MA, León Juárez M. COVID-19, una emergencia de salud pública mundial. Rev Clínica Esp [Internet]. 20 de marzo de 2020 [citado 4 de mayo de 2020]; Disponible en: <http://www.sciencedirect.com/science/article/pii/S0014256520300928>
2. Johns Hopkins Coronavirus Resource Center. COVID-19 Map [Internet]. Johns Hopkins Coronavirus Resource Center. [citado 4 de mayo de 2020]. Disponible en: <https://coronavirus.jhu.edu/map.html>
3. Médicos con Covid-19 Positivo (Autoreporte) [Internet]. Colegio Médico del Perú - Consejo Nacional. [citado 30 de junio de 2020]. Disponible en: <https://www.cmp.org.pe/medicos-con-covid-19-positivo-autoreporte/>
4. Yao X, Ye F, Zhang M, Cui C, Huang B, Niu P, et al. In Vitro Antiviral Activity and Projection of Optimized Dosing Design of Hydroxychloroquine for the Treatment of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). Clin Infect Dis Off Publ Infect Dis Soc Am. 9 de marzo de 2020;
5. National Library of Medicine, National Institutes of Health. Home - ClinicalTrials.gov [Internet]. [citado 8 de junio de 2020]. Disponible en: <https://clinicaltrials.gov/ct2/home>

6. Randomised Evaluation of COVID-19 Therapy - Full Text View - ClinicalTrials.gov [Internet]. [citado 1 de julio de 2020]. Disponible en: <https://clinicaltrials.gov/ct2/show/NCT04381936>

7. ISRCTN - ISRCTN50189673: A randomised trial of treatments to prevent death in patients hospitalised with COVID-19 (coronavirus) [Internet]. [citado 1 de julio de 2020]. Disponible en: <http://www.isrctn.com/ISRCTN50189673>

8. Soto A, Quiñones-Laveriano DM, García PJ, Gotuzzo E, Henao-Restrepo AM. Respuestas rápidas a la pandemia de COVID-19 a través de la ciencia y la colaboración global: el ensayo clínico Solidaridad. Rev Peru Med Exp Salud Pública [Internet]. 18 de mayo de 2020 [citado 8 de junio de 2020];37(2). Disponible en: <https://rpmesp.ins.gob.pe/index.php/rpmesp/article/view/5546>

9. “Solidarity” clinical trial for COVID-19 treatments [Internet]. [citado 8 de junio de 2020]. Disponible en: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/solidarity-clinical-trial-for-covid-19-treatments>

10. Universidad Peruana Cayetano Heredia. Hydroxychloroquine to Prevent SARS-CoV-2 Infection Among Healthcare Workers: Randomized Controlled, Open-label, Phase 3 Clinical Trial [Internet]. clinicaltrials.gov; 2020 jun [citado 4 de junio de 2020]. Report No.: NCT04414241. Disponible en: <https://clinicaltrials.gov/ct2/show/NCT04414241>

11. Allison D. Chatbots in the library: is it time? *Libr Hi Tech.* 2 de marzo de 2012;30(1):95-107.
12. MINSA. Covid 19 en el Perú - Ministerio del Salud [Internet]. [citado 18 de enero de 2021]. Disponible en: https://covid19.minsa.gob.pe/sala_situacional.asp
13. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol.* 2 de marzo de 2020;1-9.
14. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol.* 29 de 2016;3(1):237-61.
15. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell.* 16 de abril de 2020;181(2):271-280.e8.
16. Rodriguez-Morales AJ, Sánchez-Duque JA, Hernández Botero S, Pérez-Díaz CE, Villamil-Gómez WE, Méndez CA, et al. Preparación y control de la enfermedad por coronavirus 2019 (COVID-19) en América Latina. *ACTA MEDICA Peru.* 18 de marzo de 2020;37(1):3-7.
17. Guo Y-R, Cao Q-D, Hong Z-S, Tan Y-Y, Chen S-D, Jin H-J, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. *Mil Med Res.* 13 de marzo de 2020;7(1):11.

18. Trilla A. Un mundo, una salud: la epidemia por el nuevo coronavirus COVID-19. *Med Clin (Barc)*. 13 de marzo de 2020;154(5):175-7.
19. Cruz-Vargas JADL. Protegiendo al personal de la salud en la pandemia Covid-19. *Rev Fac Med Humana*. 22 de marzo de 2020;20(2):3-3.
20. World Health Organization. Medidas de protección básicas contra el nuevo coronavirus [Internet]. 2020 [citado 1 de mayo de 2020]. Disponible en: <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019/advice-for-public>
21. Villegas-Chiroque M. Pandemia de COVID-19: pelea o huye. *Rev Exp En Med Hosp Reg Lambayeque*. 2020;6(1).
22. Wang J, Zhou M, Liu F. Reasons for healthcare workers becoming infected with novel coronavirus disease 2019 (COVID-19) in China. *J Hosp Infect*. 6 de marzo de 2020;
23. Kouznetsov VV, Torres DFA. Antimaláricos: construcción de híbridos moleculares de la cloroquina. *Univ Sci*. 2008;13:15.
24. Hahnemann Samuel. *Organon de la medicina: el libro fundamental de la homeopatía* escrito por su fundador. México: Porrúa; 2004.
25. Bhattacharjee MK. *Chemistry of Antibiotics and Related Drugs*. Springer; 2016. 227 p.
26. Touret F, de Lamballerie X. Of chloroquine and COVID-19. *Antiviral Res*. 1 de mayo de 2020;177:104762.

27. Younis NK, Zareef RO, Al Hassan SN, Bitar F, Eid AH, Arabi M. Hydroxychloroquine in COVID-19 Patients: Pros and Cons. *Front Pharmacol* [Internet]. 2020 [citado 31 de diciembre de 2020];11. Disponible en: <https://www.frontiersin.org/articles/10.3389/fphar.2020.597985/full>
28. Chen P-L, Lee N-Y, Cia C-T, Ko W-C, Hsueh P-R. A Review of Treatment of Coronavirus Disease 2019 (COVID-19): Therapeutic Repurposing and Unmet Clinical Needs. *Front Pharmacol* [Internet]. 2020 [citado 31 de diciembre de 2020];11. Disponible en: <https://www.frontiersin.org/articles/10.3389/fphar.2020.584956/full>
29. Monti M, Vertogen B, Masini C, Donati C, Lilli C, Zingaretti C, et al. Hydroxychloroquine as Prophylaxis for COVID-19: A Review. *Front Pharmacol* [Internet]. 2020 [citado 31 de diciembre de 2020];11. Disponible en: <https://www.frontiersin.org/articles/10.3389/fphar.2020.605185/full>
30. Rich E. *Artificial Intelligence*. New York, NY, USA: McGraw-Hill, Inc.; 1983.
31. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. *Npj Digit Med*. 14 de marzo de 2018;1(1):1-4.
32. Turing AM. Computing Machinery and Intelligence. *Mind*. 1 de octubre de 1950;LIX(236):433-60.
33. Rosenblatt F. *The perceptron, a perceiving and recognizing automaton*. Buffalo, NY: Cornell Aeronautical Laboratory; 1957.

34. Daniel B. Natural Language Input for a Computer Problem Solving System. 1 de marzo de 1964; Disponible en: <http://hdl.handle.net/1721.1/5922>
35. Smith RG. Knowledge-Based Systems: Concepts, Techniques, Examples. 8 de mayo de 1985;84.
36. Franklin S., Graesser A., ECAI Workshop on Agent Theories A and Languages, ATAL 1996. Is It an agent, or just a program?: A taxonomy for autonomous agents. Lect Notes Comput Sci Lect Notes Comput Sci Subser Lect Notes Artif Intell Lect Notes Bioinforma. 1997;1193:21-35.
37. Ferrucci D, Levas A, Bagchi S, Gondek D, Mueller ET. Watson: Beyond Jeopardy! Artif Intell. 1 de junio de 2013;199-200:93-105.
38. Luger GF. Artificial Intelligence: Structures and Strategies for Complex Problem Solving. 6th ed. USA: Addison-Wesley Publishing Company; 2008.
39. Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 3 edition. Upper Saddle River: Pearson; 2009. 1152 p.
40. Poole DL, Mackworth AK, Goebel R. Computational intelligence: a logical approach [Internet]. New York: Oxford University Press; 1998 [citado 29 de octubre de 2019]. Disponible en: <https://archive.org/details/computationalint00pool>
41. Goel B. Developments in The Field of Natural Language Processing. Int J Adv Res Comput Sci. 2017;7.

42. Brassai ST, Iantovics B, Enăchescu C. Artificial Intelligence in the Path Planning Optimization of Mobile Agent Navigation. *Procedia Econ Finance*. 1 de enero de 2012;3:243-50.
43. Scassellati B. Theory of Mind for a Humanoid Robot. *Auton Robots*. 1 de enero de 2002;12(1):13-24.
44. Kleene SC. Representation of Events in Nerve Nets and Finite Automata. En: *Automata Studies* [Internet]. Princeton: Princeton University Press; 1956 [citado 30 de octubre de 2019]. Disponible en: doi.org/10.1515/9781400882618
45. Freund Y, Schapire RE. Large Margin Classification Using the Perceptron Algorithm. *Mach Learn*. 1 de diciembre de 1999;37(3):277-96.
46. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Netw*. 1 de enero de 2015;61:85-117.
47. Kong L, Dyer C, Smith NA. Segmental Recurrent Neural Networks. 1 de marzo de 2016 [citado 2 de diciembre de 2019]; Disponible en: <http://arxiv.org/abs/1511.06018>
48. Su M-H, Wu C-H, Huang K-Y, Hong Q-B, Wang H-M. A chatbot using LSTM-based multi-layer embedding for elderly care. En: *2017 International Conference on Orange Technologies (ICOT)*. 2017. p. 70-4.
49. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. octubre de 1986;323(6088):533-6.

50. Accenture Interactive. Chatbots in customer Service [Internet]. Accenture Digital; 2016. Disponible en: https://www.accenture.com/t00010101T000000__w__/br-pt/_acnmedia/PDF-45/Accenture-Chatbots-Customer-Service.pdf

51. Janarthanam S. Hands On Chatbots and Conversational UI Development. Vol. 1. Packt; 2017. 111 p.

52. Warwick K, Shah H. Good Machine Performance in Turing's Imitation Game. IEEE Trans Comput Intell AI Games. septiembre de 2014;6(3):289-99.

53. Radziwill N, Benton M. Evaluating Quality of Chatbots and Intelligent Conversational Agents. :21.

54. AI Multiple. 15 Key Metrics for Chatbot Conversational Analytics in 2020 [Internet]. AI Multiple. 2020 [citado 10 de enero de 2020]. Disponible en: <https://blog.aimultiple.com/chatbot-analytics/>

55. U.S. Department of Health & Human Services. Usability Evaluation Basics [Internet]. Department of Health and Human Services; 2013 [citado 29 de junio de 2020]. Disponible en: [usability-evaluation.html](#)

56. ISO 9241-11. ISO 9241-11:1998(en), Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability [Internet]. [citado 29 de junio de 2020]. Disponible en: <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en>

57. Lewis JR. Measuring Perceived Usability: The CSUQ, SUS, and UMUX. *Int J Human-Computer Interact.* 2 de diciembre de 2018;34(12):1148-56.
58. Buttle F, Maklan S. *Customer Relationship Management: Concepts and Technologies.* Routledge; 2019. 469 p.
59. SAP. About SAP SE [Internet]. SAP. [citado 2 de mayo de 2020]. Disponible en: <https://www.sap.com/corporate/en.html>
60. Salesforce. Salesforce.com: The Customer Success Platform To Grow Your Business [Internet]. Salesforce.com. [citado 2 de mayo de 2020]. Disponible en: <https://www.salesforce.com/mx/>
61. Sirena. The first CRM integrated to WhatsApp · Sirena [Internet]. [citado 3 de mayo de 2020]. Disponible en: <https://www.sirena.app/en-us/>
62. Zoho. Zoho - Cloud Software Suite and SaaS Applications for Businesses [Internet]. Zoho. [citado 2 de mayo de 2020]. Disponible en: <https://www.zoho.com/index.html>
63. Aguilar MIH, Villegas AAG. Análisis comparativo de la Escala de Usabilidad del Sistema (EUS) en dos versiones / Comparative analysis of the System Usability Scale (SUS) in two versions. *RECI Rev Iberoam Las Cienc Comput E Informática.* 11 de julio de 2016;5(10):44-58.
64. Brooke J. SUS - A quick and dirty usability scale. Redhatch Consult Ltd. 1996;7.

65. Sauro J, Lewis JR. When designing usability questionnaires, does it hurt to be positive? En: Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11 [Internet]. Vancouver, BC, Canada: ACM Press; 2011 [citado 29 de junio de 2020]. p. 2215. Disponible en: <http://dl.acm.org/citation.cfm?doid=1978942.1979266>
66. Sauro J. MeasuringU: 5 Ways to Interpret a SUS Score [Internet]. MeasuringU: 5 Ways to Interpret a SUS Score. 2018 [citado 30 de junio de 2020]. Disponible en: <https://measuringu.com/interpret-sus-score/>
67. Bangor A. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. 2009;4(3):10.
68. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. Int J Hum-Comput Interact. 29 de julio de 2008;24(6):574-94.
69. MeasuringU: Predicting Net Promoter Scores from System Usability Scale Scores [Internet]. [citado 7 de febrero de 2021]. Disponible en: <https://measuringu.com/nps-sus/>
70. Fred D. D. User Acceptance of Information Technology: system characteristics, user perceptions and behavioral impacts. Man-Machine Studies [Internet]. 8 de diciembre de 1991;38. Disponible en: <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/30954/0000626.pdf;jsessionid=56E6EC4AEAF219CD1D0C100B416F99FE?sequence=1>

71. Epi Info™ | CDC [Internet]. 2019 [citado 1 de enero de 2021]. Disponible en: https://www.cdc.gov/epiinfo/esp/es_index.html
72. RStudio | Open source & professional software for data science teams [Internet]. [citado 15 de julio de 2021]. Disponible en: <https://rstudio.com/>
73. Barret E. How a chatbot helped secure \$10million for Ovarian Cancer Canada [Internet]. 2019. Disponible en: <https://www.maybetech.com/blog/case-study/how-a-chatbot-helped-secure-a-10million-investment/>
74. Your.MD Ltd. Your.MD - Health Guide and Self-Care Checker [Internet]. Your.MD. [citado 8 de enero de 2020]. Disponible en: <https://www.your.md/>
75. Ada: Your health companion [Internet]. Ada. [citado 30 de enero de 2020]. Disponible en: <https://ada.com/>
76. Sadavarte SS, Bodanese E. Pregnancy Companion Chatbot Using Alexa and Amazon Web Services. En: 2019 IEEE Pune Section International Conference (PuneCon) [Internet]. Pune, India: IEEE; 2019 [citado 1 de mayo de 2021]. p. 1-5. Disponible en: <https://ieeexplore.ieee.org/document/9105762/>
77. Pregnancy, baby and toddler health information at BabyCentre UK - BabyCentre UK [Internet]. [citado 8 de enero de 2020]. Disponible en: <https://www.babycentre.co.uk/>
78. Chaix B, Bibault J-E, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, et al. When Chatbots Meet Patients: One-Year Prospective Study of

Conversations Between Patients With Breast Cancer and a Chatbot. *JMIR Cancer*. 2 de mayo de 2019;5(1):e12856.

79. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health*. 6 de junio de 2017;4(2):e19.
80. Recopilación de enfoques de gobierno abierto para COVID-19 [Internet]. Open Government Partnership. [citado 27 de noviembre de 2020]. Disponible en: <https://www.opengovpartnership.org/es/collecting-open-government-approaches-to-covid-19/>
81. OPS. EL POTENCIAL DE LAS TECNOLOGÍAS DE LA INFORMACIÓN DE USO FRECUENTE DURANTE LA PANDEMIA [Internet]. DEPARTAMENTO DE EVIDENCIA E INTELIGENCIA PARA LA ACCIÓN EN SALUD; 2020. Disponible en: <https://iris.paho.org/handle/10665.2/52022>
82. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. How the COVID-19 pandemic is favoring the adoption of digital technologies in healthcare: a literature review. *medRxiv*. 18 de mayo de 2020;2020.04.26.20080341.
83. Chatbots: nuevo aliado para combatir al COVID-19 | Carmelo Portal [Internet]. 2020 [citado 27 de noviembre de 2020]. Disponible en:

<https://www.carmeloportal.com/54801-chatbots-nuevo-aliado-para-combatir-al-covid-19>

84. Bots HN. El detrás de escena de la estrategia digital del Plan Coronavirus [Internet]. heynow. 2020 [citado 27 de noviembre de 2020]. Disponible en: <https://www.heynowbots.com/post/el-detrás-de-escena-de-la-estrategia-digital-del-plan-coronavirus>
85. El secreto del éxito del Uruguay contra la Covid-19 [Internet]. [citado 27 de noviembre de 2020]. Disponible en: http://www.anp.org.pe/index.php?option=com_content&view=article&id=3543&Itemid=97
86. ElPais. Habilitan chatbot de dudas por coronavirus: cómo funciona y qué se puede preguntar [Internet]. Diario EL PAIS Uruguay. [citado 27 de noviembre de 2020]. Disponible en: <https://www.elpais.com.uy/informacion/salud/habilitan-chatbot-dudas-coronavirus-funciona-preguntar.html>
87. Ministerio de Salud Pública [Internet]. Ministerio de Salud Pública. [citado 27 de noviembre de 2020]. Disponible en: <https://www.gub.uy/ministerio-salud-publica/coronavirus>
88. Buenos Aires uses WhatsApp to assist COVID-19 response [Internet]. Cities Today - Connecting the world's urban leaders. 2020 [citado 27 de noviembre de 2020]. Disponible en: <https://cities-today.com/buenos-aires-uses-whatsapp-to-assist-covid-19-response/>

89. El chatbot Boti identifica 3.000 casos de Covid-19 en Buenos Aires. [Internet]. Parada Visual. 2020 [citado 27 de noviembre de 2020]. Disponible en: <https://www.paradavisual.com/un-chatbot-identifica-3-000-casos-de-covid-19-en-buenos-aires/>
90. El chat porteño ya recibió más de 40.000 consultas por coronavirus | Noticias | Buenos Aires Ciudad - Gobierno de la Ciudad Autónoma de Buenos Aires [Internet]. [citado 27 de noviembre de 2020]. Disponible en: <https://www.buenosaires.gob.ar/laciudad/noticias/el-chat-de-la-Ciudad-ya-recibio-mas-de-35000-consultas-por-sintomas-de-coronavirus>
91. Buenos Aires chatbot helps to protect citizens against Coronavirus | Cities for Global Health [Internet]. [citado 27 de noviembre de 2020]. Disponible en: <https://www.citiesforglobalhealth.org/initiative/buenos-aires-chatbot-helps-protect-citizens-against-coronavirus>
92. El gobierno porteño lanzó un chatbot de WhatsApp para el seguimiento de contactos estrechos de casos de coronavirus [Internet]. Infobae. [citado 27 de noviembre de 2020]. Disponible en: <https://www.infobae.com/tecno/2020/08/06/el-gobierno-porteno-lanzo-un-chatbot-de-whatsapp-para-el-seguimiento-de-contactos-estrechos-de-casos-de-coronavirus/>
93. Ministerio de Salud de la Provincia del Neuquén [Internet]. [citado 27 de noviembre de 2020]. Disponible en: <https://www.saludneuquen.gob.ar/conoce-a-andy-el-robot-del-ministerio-de-salud/>

94. Tecnología argentina para combatir infodemia y fake news sobre COVID-19 [Internet]. ITSitio. 2020 [citado 27 de noviembre de 2020]. Disponible en: <https://www.itsitio.com/ar/tecnologia-argentina-combatir-infodemia-fake-news-covid-19/>
95. Salud Digital | Servicios de Salud en La Paz, Bolivia presentaron una chatbot de consultas y diagnósticos en línea para personas con síntomas de COVID-19 [Internet]. Salud Digital. 2020 [citado 27 de noviembre de 2020]. Disponible en: <https://saluddigital.com/plataformas-digitales/servicios-de-salud-en-la-paz-bolivia-presentaron-una-chatbot-de-consultas-y-diagnosticos-en-linea-para-personas-con-sintomas-de-covid-19/>
96. Rivera A. El chatbot inteligente que detecta el Covid-19 se abre paso entre las compañías mineras [Internet]. Diario Financiero. [citado 27 de noviembre de 2020]. Disponible en: <http://www.df.cl/noticias/df-lab/transformacion-digital/el-chatbot-inteligente-que-detecta-el-covid-19-se-abre-paso-entre-las/2020-04-21/190133.html>
97. 1MillionBot - Chatbot Catalina - Ecuador COVID-19 [Internet]. 1MillionBot. [citado 27 de noviembre de 2020]. Disponible en: <https://1millionbot.com/chatbot-coronavirus-ecuador/>
98. Munsch N, Martin A, Gruarin S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, et al. A benchmark of online COVID-19 symptom checkers. medRxiv. 19 de agosto de 2020;2020.05.22.20109777.

99. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. medRxiv. 8 de julio de 2020;2020.07.07.20147975.
100. Battineni G, Chintalapudi N, Amenta F. AI Chatbot Design during an Epidemic like the Novel Coronavirus. Healthcare. junio de 2020;8(2):154.
101. Martin A, Nateqi J, Gruarin S, Munsch N, Abdarahmane I, Zobel M, et al. An artificial intelligence-based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot. Sci Rep. 4 de noviembre de 2020;10(1):19012.
102. Mehl A, Bergey F, Cawley C, Gilsdorf A. Syndromic surveillance insights from a symptom assessment app before and during COVID-19 measures in Germany and the United Kingdom: results from repeated cross-sectional analyses. medRxiv. 19 de junio de 2020;2020.06.16.20126466.
103. Oniani D, Wang Y. A Qualitative Evaluation of Language Models on Automatic Question-Answering for COVID-19. ArXiv200610964 Cs [Internet]. 23 de junio de 2020 [citado 28 de noviembre de 2020]; Disponible en: <http://arxiv.org/abs/2006.10964>
104. Sermet Y, Demir I. A Semantic Web Framework for Automated Smart Assistants: COVID-19 Case Study. ArXiv200700747 Cs [Internet]. 17 de septiembre de 2020 [citado 28 de noviembre de 2020]; Disponible en: <http://arxiv.org/abs/2007.00747>

105. Hwerbi K. An ontology-based chatbot for crises management: use case coronavirus. ArXiv201102340 Cs [Internet]. 2 de noviembre de 2020 [citado 27 de noviembre de 2020]; Disponible en: <http://arxiv.org/abs/2011.02340>
106. HerrimanMaguire, MeerElana, RosinRoy, LeeVivian, WashingtonVindell, G V. Asked and Answered: Building a Chatbot to Address Covid-19-Related Concerns. NEJM Catal Innov Care Deliv [Internet]. 18 de junio de 2020 [citado 27 de noviembre de 2020]; Disponible en: <https://catalyst.nejm.org/doi/full/10.1056/cat.20.0230>
107. Yang W, Zeng G, Tan B, Ju Z, Chakravorty S, He X, et al. On the Generation of Medical Dialogues for COVID-19. medRxiv. 15 de mayo de 2020;2020.05.08.20095810.
108. Khedlekar SK, Suryawanshi VR, Khedlekar SK. Problem Solving Chatbot for Covid-19. J Inf Comput Sci. 2020;10(6):11.
109. Ouerhani N, Maalel A, Ghézala HB, Chouri S. Smart Ubiquitous Chatbot for COVID-19 Assistance with Deep learning Sentiment Analysis Model during and after quarantine [Internet]. In Review; 2020 jun [citado 27 de noviembre de 2020]. Disponible en: <https://www.researchsquare.com/article/rs-33343/v1>
110. Li Y, Grandison T, Silveyra P, Douraghy A, Guan X, Kieselbach T, et al. Jennifer for COVID-19: An NLP-Powered Chatbot Built for the People and by the People to Combat Misinformation. 23 de junio de 2020 [citado 27 de noviembre de 2020]; Disponible en: https://openreview.net/forum?id=HxIZzQZy_0F

111. Pandey R, Gautam V, Pal R, Bandhey H, Singh L, Sharma H, et al. A Machine Learning Application for Raising WASH Awareness in the Times of COVID-19 Pandemic. :14.
112. Judson TJ, Odisho AY, Young JJ, Bigazzi O, Steuer D, Gonzales R, et al. Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. *J Am Med Inform Assoc.* 1 de septiembre de 2020;27(9):1450-5.
113. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ.* 7 de abril de 2020;m1328.
114. Turk P, Tran T, Rose G, McWilliams A. A Predictive Internet-Based Model for COVID-19 Hospitalization Census. *medRxiv.* 18 de noviembre de 2020;2020.11.15.20231845.
115. Maniou TA, Veglis A. Employing a Chatbot for News Dissemination during Crisis: Design, Implementation and Evaluation. *Future Internet.* julio de 2020;12(7):109.
116. Kadariya D, Venkataramanan R, Yip HY, Kalra M, Thirunarayanan K, Sheth A. kBot: Knowledge-enabled Personalized Chatbot for Asthma Self-Management. *Proc Int Conf Smart Comput SMARTCOMP Int Conf Smart Comput.* junio de 2019;2019:138-43.
117. Tielman ML, Neerincx MA, Bidarra R, Kybartas B, Brinkman W-P. A Therapy System for Post-Traumatic Stress Disorder Using a Virtual Agent and

- Virtual Storytelling to Reconstruct Traumatic Memories. *J Med Syst* [Internet]. 2017 [citado 7 de abril de 2021];41(8). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5506234/>
118. Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A. Usability, Acceptability, and Effectiveness of Web-Based Conversational Agents to Facilitate Problem Solving in Older Adults: Controlled Study. *J Med Internet Res*. 27 de 2020;22(5):e16794.
119. Waselewski ME, Flickinger TE, Canan C, Harrington W, Franklin T, Otero KN, et al. A Mobile Health App to Support Patients Receiving Medication-Assisted Treatment for Opioid Use Disorder: Development and Feasibility Study. *JMIR Form Res* [Internet]. 23 de febrero de 2021 [citado 21 de marzo de 2021];5(2). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7943342/>
120. Djalev L, Bogdanov S. Age and Gender Differences in Evaluating the Pedagogical Usability of E-Learning Materials. *Engl Stud NBU*. 30 de diciembre de 2019;5(2):169-89.
121. Maudlin LC, McNeal KS, Dinon-Aldridge H, Davis C, Boyles R, Atkins RM. Website Usability Differences between Males and Females: An Eye-Tracking Evaluation of a Climate Decision Support System. *Weather Clim Soc*. enero de 2020;12(1):183-92.
122. Bickmore TW, Trinh H, Olafsson S, O’Leary TK, Asadi R, Rickles NM, et al. Patient and Consumer Safety Risks When Using Conversational Assistants for

Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. *J Med Internet Res.* 4 de septiembre de 2018;20(9):e11510.

123. Kowatsch T, Volland D, Shih I, Rügger D, Künzler F, Barata F, et al. Design and Evaluation of a Mobile Chat App for the Open Source Behavioral Health Intervention Platform MobileCoach. En: *Designing the Digital Transformation* [Internet]. Springer, Cham; 2017 [citado 12 de abril de 2021]. p. 485-9. Disponible en: https://link.springer.com/chapter/10.1007/978-3-319-59144-5_36
124. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, O'Donnell J. Response to a Relational Agent by Hospital Patients with Depressive Symptoms. *Interact Comput.* 1 de julio de 2010;22(4):289-98.
125. McKenna KC, Geoghegan C, Swezey T, Perry B, Wood WA, Nido V, et al. Investigator Experiences Using Mobile Technologies in Clinical Research: Qualitative Descriptive Study. *JMIR MHealth UHealth* [Internet]. 12 de febrero de 2021 [citado 21 de marzo de 2021];9(2). Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7910119/>
126. Kocielnik R, Xiao L, Avrahami D, Hsieh G. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proc ACM Interact Mob Wearable Ubiquitous Technol.* 5 de julio de 2018;2(2):1-26.
127. Mendoza V, Novick DG. Usability over time. En: *Proceedings of the 23rd annual international conference on Design of communication documenting &*

designing for pervasive information - SIGDOC '05 [Internet]. Coventry, United Kingdom: ACM Press; 2005 [citado 2 de mayo de 2021]. p. 151. Disponible en: <http://portal.acm.org/citation.cfm?doid=1085313.1085348>

128. Kjeldskov J, Skov MB, Stage J. A longitudinal study of usability in health care: Does time heal? *Int J Med Inf.* junio de 2010;79(6):e135-43.

ANEXO I

