

Towards a Set of Heuristics for Evaluating Chatbots

L. Sánchez, S. Mendoza, J. Urquiza, J. Rodríguez, and A. Meneses

Abstract—Chatbots are artificial intelligence tools that interact with people in different contexts. A chatbot can be useful to streamline daily processes, serve customers 24 hours a day, provide information about classes, among other things. The appearance of new development technologies has made creating a chatbot an increasingly fast and straightforward process, bringing this kind of applications to people who had never considered using them before. However, this speed in development can lead to specific problems, many of them caused by the lack of usability evaluations. Heuristic usability evaluations are user interface review processes carried out by experts and are an essential part of any assessment process. To date, there are no heuristics to evaluate the usability of chatbots. Therefore, this work proposes five usability heuristics in chatbots that come from the experience developing this type of applications, as well as from a broad review of state of the art. The set of heuristics was tested using a case study with the help of five experts, who evaluated an education-oriented chatbot. The results revealed that, although the proposed heuristics need refinement, they are an excellent first step in broadening the horizon of usability evaluations in chatbots.

Index Terms—Chatbots, Evaluation, Heuristics, Usability.

I. INTRODUCCIÓN

En los últimos años, se ha observado un interés cada vez mayor por los chatbots, que son programas que usan procesamiento del lenguaje natural (NLP por sus siglas en inglés) para interactuar con humanos bajo diferentes contextos [1]. De acuerdo a *Business Insider*, se espera que, en 2024, los consumidores gasten hasta 142 mil millones de dólares mediante este tipo de sistemas [2]. Los chatbots se pueden encontrar en diversos entornos, e.g., negocios [3], turismo [4], preguntas frecuentes [5], procedimientos [6] y recomendaciones [7]. Usualmente, los chatbots funcionan mediante la búsqueda de palabras clave, frases o ejemplos que tengan almacenados en sus bases de conocimientos, con el objetivo de ofrecer información acerca de productos o servicios, actividades o lugares, dentro de redes sociales o sitios web [5].

El distanciamiento social, causado por la pandemia del SARS-CoV-2, ha puesto en evidencia la importancia vital que tienen los medios de comunicación digitales [8]. De esta manera, los chatbots han cobrado una notoriedad considerable, pues se han desarrollado como herramientas para combatir la

propia pandemia [9], para brindar atención al cliente [10], así como en muchos otros ámbitos comerciales y gubernamentales [11]. Esta rampante demanda de chatbots no sólo se debe a que son una herramienta que puede ofrecer comunicación inmediata y automatización de ciertas tareas, sino también a las diferentes tecnologías que permiten desarrollos rápidos, pues se pueden obtener chatbots funcionales en cuestión de horas [12].

Los desarrollos rápidos son valiosos, ya que permiten a los desarrolladores obtener diferentes puntos de vista [13]. Sin embargo, en el caso de los chatbots (y otras tecnologías) esta premura puede crear más problemas de los que se intentan solucionar porque los chatbots pueden ser lanzados directamente a los usuarios finales con evaluaciones de usabilidad deficientes o nulas. En consecuencia, la interacción se dificulta y el problema original se agudiza, puesto que todo el proceso de comunicación se vuelve más pobre y deficiente [14]. Es evidente entonces que las evaluaciones de usabilidad son necesarias.

La definición más conocida de usabilidad es propuesta por la ISO (9241-11:2018): «La medida en que un producto puede ser utilizado por usuarios específicos para lograr objetivos específicos con eficacia, eficiencia y satisfacción en un contexto de uso específico». Esta definición abarca tres ejes principales: **1) usuarios específicos:** no cualquier usuario, sino el usuario específico para el que el producto fue diseñado; **2) objetivos específicos:** los usuarios específicos tienen que compartir un objetivo para el producto, i.e., que su finalidad sea el objetivo del producto; **3) contexto de uso específico:** el producto tiene que estar diseñado para trabajar en el ambiente en donde los usuarios lo utilizarán [15]. La usabilidad es parte esencial de la experiencia de usuario (UX por sus siglas en inglés) y son conceptos complementarios, mas nunca intercambiables [16].

Definir la UX es un asunto complejo, ya que no existe una explicación uniforme que comprenda todos los contextos [17], [18]. No obstante, se puede hacer uso nuevamente de la ISO (9241-11:2018) que da un buen concepto general de UX: «Las percepciones y respuestas de una persona que resultan del uso o uso anticipado de un producto, sistema o servicio». Sin importar la definición, todos los expertos están de acuerdo en que la usabilidad es un elemento vital para la UX, e.g., el *panal de abeja* de Morville [19] considera a la usabilidad como parte de la UX.

La importancia de las evaluaciones de usabilidad y de UX ha quedado bien establecida en el estado del arte, especialmente en el ámbito educativo [20], [21], [22], [23]. El término «evaluación» es amplio, pues existen diversas herramientas y métodos para realizar evaluaciones de usabilidad. Sin embargo, el contexto de este artículo sólo se centra en evaluaciones heurísticas. Una evaluación heurística (también llamada experta) es una valoración o inspección

L. Sánchez, Departamento de Computación, CINVESTAV-IPN (luismartin.sanchez@cinvestav.mx).

S. Mendoza, Departamento de Computación, CINVESTAV-IPN (smendoza@cs.cinvestav.mx).

J. Urquiza, Departamento de Computación, CINVESTAV-IPN (fuy@computacion.cs.cinvestav.mx).

J. Rodríguez, Departamento de Computación, CINVESTAV-IPN (rodriguez@cs.cinvestav.mx).

A. Meneses, Departamento de Computación, CINVESTAV-IPN (ameneses@cs.cinvestav.mx).

de un producto hecha por expertos [24]. Típicamente esto significa expertos en usabilidad, pero también puede significar doble pericia: tanto en usabilidad, como en el dominio del producto. Las heurísticas son un conjunto de principios o reglas generales, usadas por los expertos, para inspeccionar una interfaz de usuario en búsqueda de violaciones de dichas reglas [24]. Cabe destacar que una evaluación heurística, al igual que cualquier otro método individual, no representa una evaluación de usabilidad completa, i.e., hacen falta diversas herramientas en diferentes etapas del desarrollo para tener una valoración integral. Por lo general, las evaluaciones heurísticas representan la primera fase de pruebas de un producto o servicio [24].

De esta manera, se decidió enfocarse en proponer heurísticas de usabilidad para chatbots. En primer lugar, porque no se encontraron heurísticas similares en la literatura científica. En segundo lugar, porque el número de sistemas que dependen de la inteligencia artificial ha ido en aumento y muchas veces estos se tratan como cajas negras, por lo que también requieren comprensión y mecanismos de evaluación por parte de la comunidad de UX [25]. Finalmente, porque un conjunto de heurísticas puede adaptarse al desarrollo frenético de los chatbots hoy en día y representa una inversión mínima, que puede traer grandes beneficios. Además, la usabilidad es una parte importante para que los usuarios tengan claro el uso de sus datos, especialmente en sistemas que buscan ayudar en medio de una pandemia [26].

Se proponen cinco heurísticas para evaluar la usabilidad de los chatbots y para probarlas, se reunieron a cinco expertos en usabilidad y UX para que evaluaran un chatbot centrado en la educación.

A continuación se describe la organización de este artículo. Primero, se analizan algunos trabajos relacionados al tópico de estudio (ver Sección II). Después, se describe la metodología que se siguió para la elaboración de la propuesta (ver Sección III). Seguidamente, se explica el desarrollo de las heurísticas de usabilidad para chatbots (ver Sección IV). Más adelante, se detalla el caso de estudio (ver Sección V). Finalmente, se cierra con las conclusiones y algunas ideas para trabajo futuro (ver Sección VI).

II. TRABAJOS RELACIONADOS

En esta sección se presentan algunos trabajos que demuestran que la evaluación de chatbots es un tópico valioso, así como la importancia de la propuesta descrita en este artículo.

Valtolina et al. [27] presentaron un estudio que resalta los beneficios de desarrollar un modelo conceptual basado en un estilo de interacción conversacional, con el fin de permitir a los usuarios comunicarse con un sistema de una manera que les resulte familiar. Si bien evaluaron diversas aplicaciones en los ámbitos de salud y domótica e identificaron problemas abiertos, su análisis es relativamente superficial, pues utilizaron cuestionarios de propósito general bien conocidos en el estado del arte.

Kocaballi et al. [28] estudiaron una variedad de cuestionarios de UX para valorar su cobertura al evaluar chatbots. En primer lugar, descubrieron que muchos estudios utilizaban los

conceptos de UX y usabilidad indistintamente, lo que sesgaba los resultados. Después, estudiaron cuántas y cuáles dimensiones cubrían los cuestionarios. Concluyeron que, aunque se utilicen varios cuestionarios para una misma evaluación de un chatbot, no es suficiente para realizar una evaluación completa, pues no se identificaron todas las dimensiones relevantes para los sistemas conversacionales.

Ren et al. [29] realizaron una revisión del estado del arte que versa sobre las técnicas de evaluación de usabilidad en chatbots. Identificaron diversos elementos a considerar para una evaluación satisfactoria dentro de los componentes clásicos de la usabilidad: efectividad (e.g., completitud de tareas y certeza), eficiencia (e.g., tiempo para completar una tarea y esfuerzo mental) y satisfacción (e.g., facilidad de uso y preguntas dependientes del contexto).

Ding et al. [30] desarrollaron un panel en el que se discutió cómo los medios de interacción tradicionales (e.g., sitios web) están migrando en forma de chatbots y todo lo que esto representa, e.g., investigación centrada en el usuario, pruebas de usabilidad, arquitectura de la información y competencias técnicas de desarrollo. Los autores sostienen que una evaluación heurística es una herramienta simple pero efectiva para realizar valoraciones de usabilidad en chatbots.

Holmes et al. [31] evaluaron, con una diversidad herramientas, un chatbot que ayuda con el control de peso. Mediante la aplicación de dos cuestionarios, entrevistas y tareas, detallaron el nivel de convergencia de los cuestionarios, concluyendo que para evaluar la usabilidad de un chatbot se requieren métodos no tradicionales, puesto que los cuestionarios no tomaron en cuenta todas las posibles dimensiones a evaluar.

Yin et al. [32] investigaron el impacto que tienen los chatbots en el aprendizaje, la motivación y el rendimiento de alumnos universitarios. Evaluaron, en una sesión, a dos grupos de una clase de computación básica: el primero tuvo una sesión tradicional, i.e., una clase cara a cara y el segundo grupo utilizó un chatbot. Los resultados mostraron que ambos grupos obtuvieron un rendimiento similar, demostrando que los chatbots pueden integrarse en una clase, dado que ayudan a la motivación y al rendimiento de los estudiantes.

Rafael et al. [33] condujeron un estudio que involucraba a dos grupos de estudiantes del sistema tributario chileno. En una sesión de estudio, un grupo realizó actividades de aprendizaje activo, mientras que el otro grupo utilizó un chatbot. Gracias a los resultados de un cuestionario de preguntas abiertas, que aplicaron antes y después de la sesión, descubrieron que el grupo que utilizó el chatbot obtuvo un mejor rendimiento.

De los trabajos anteriores se puede concluir que las evaluaciones de usabilidad en chatbots son necesarias, pero no existen métodos de evaluación que se enfoquen en este tipo de sistemas. Adicionalmente, los chatbots pueden tener un impacto positivo en el proceso de enseñanza/aprendizaje. De esta manera, la propuesta de este artículo resalta de las anteriores por ofrecer una herramienta de evaluación de usabilidad centrada en chatbots, más aún porque se analizan chatbots enfocados en la educación.

III. METODOLOGÍA DE INVESTIGACIÓN

La metodología que se siguió en la creación de las heurísticas es la propuesta por Quiñones *et al.* [34]. Consta de seis etapas:

- 1) **Etapla exploratoria:** reunir bibliografía relacionada con el tema principal de la investigación.
- 2) **Etapla descriptiva:** destacar las características más importantes de la información reunida previamente.
- 3) **Etapla correlacional:** identificar las características que las heurísticas de usabilidad deben tener para el sistema particular.
- 4) **Etapla explicativa:** especificar formalmente el conjunto de heurísticas propuesto.
- 5) **Etapla de validación:** contrastar las nuevas heurísticas contra las heurísticas tradicionales por medio de experimentos.
- 6) **Etapla de refinación:** está basada en la retroalimentación obtenida en la etapa anterior.

IV. HEURÍSTICAS DE USABILIDAD PARA CHATBOTS

Esta sección contiene el desarrollo de cada una de las etapas presentadas en la metodología.

A. Etapa Exploratoria

Se llevó a cabo una búsqueda amplia de trabajos en las bibliotecas científicas digitales más reconocidas, tales como Scopus, Web of Science, Google Scholar, IEEE Xplore, ACM Digital Library, ScienceDirect y Springer Link. La Sección II de Trabajos Relacionados incluye una muestra de los artículos más relevantes para la propuesta de las heurísticas.

B. Etapa Descriptiva

Los chatbots se han usado durante décadas. El más antiguo probablemente sea «ELIZA» que data de 1966 [35]. No obstante, las bases tecnológicas de los chatbots han evolucionado rápidamente, por lo que un estándar de evaluación aún no ha sido establecido. De la búsqueda realizada en la etapa anterior, se pueden establecer tres categorías de evaluación para chatbots: 1) aquellos trabajos que se centran en los aspectos técnicos o teóricos, *i.e.*, que tienen que ver con la implementación [36] o bien con el NLP [37]; 2) las evaluaciones que involucran la satisfacción de los usuarios [38]; y 3) algunas propuestas híbridas que toman elementos de las dos primeras clasificaciones [39]. Dada la naturaleza de este artículo, únicamente se seleccionaron trabajos de la segunda categoría.

Cabe resaltar que no se encontraron heurísticas de usabilidad para chatbots, así como ninguna otra herramienta o método de evaluación diseñado específicamente para esta clase de sistemas. Sin embargo, algunas características relevantes que se encontraron son:

- Las respuestas de un chatbot se clasifican como correctas o incorrectas, de acuerdo al criterio del usuario o del evaluador [38].

- Muchos estudios presentan resultados de cuestionarios *ad hoc*, que sólo permiten clasificaciones de éxito o fracaso [40].
- Las características de usabilidad que más se miden son: eficiencia, satisfacción, efectividad, facilidad de uso, rendimiento, frustración, dificultad y esfuerzo mental [41].

C. Etapa Correlacional

Como ya se mencionó, la usabilidad clásicamente seimenta en los ejes de eficacia, eficiencia y satisfacción [42]. De acuerdo con la exploración realizada en la primera etapa y las conclusiones de la etapa descriptiva, se estima que los criterios que resumen Ren *et al.* [29], en cada eje de usabilidad, son los necesarios para evaluar chatbots:

- **Eficacia:** conclusión de tareas, precisión y recuperación.
- **Eficiencia:** tiempo para finalizar tareas, esfuerzo mental y esfuerzo comunicativo.
- **Satisfacción:** facilidad de uso, preguntas dependientes del contexto, control de complejidad, malestar físico, placer, ganas usarlo nuevamente y capacidad de aprendizaje.

Se eligió este trabajo porque: 1) recopila concretamente las características más importantes a evaluar en cada eje de usabilidad y 2) representa la visión de múltiples investigaciones, dado que es el resultado de un *survey*.

D. Etapa Explicativa

De acuerdo a los criterios mencionados en la etapa anterior, se proponen las siguientes cinco heurísticas. Para cada una se presenta la definición, ejemplos, beneficios (qué se espera cuando la heurística se cumple), así como su relación con los ejes de la usabilidad (eficacia, eficiencia y satisfacción) y la justificación de cada una, *i.e.*, la razón fundamental de su existencia con base en las características encontradas por cada eje de usabilidad:

- H1 **Completitud:** se refiere a la habilidad y flexibilidad del chatbot para entender las entradas de los usuarios y ayudarles a resolver sus problemas.
Ejemplo: si un usuario quiere crear un recordatorio, el chatbot solicita todos los datos relevantes (*e.g.*, fecha, hora y asunto).
Beneficio: crear interacciones que ayuden a los usuarios a realizar sus tareas.
Eje de usabilidad: eficacia.
Justificación: para poder finalizar las tareas del usuario con precisión, es necesario que el chatbot identifique todos los datos dependientes del contexto.
- H2 **Contexto:** indica la habilidad del chatbot para cambiar de contexto, *i.e.*, cuando conversa con un usuario ¿Qué tan fácil le es mantener la conversación? ¿Se adapta rápidamente a los cambios del usuario?
Ejemplo: si en una conversación el usuario usó la palabra «tareas» y después utilizó la palabra «deberes» bajo el mismo contexto, el chatbot se adapta.
Beneficio: ofrecer al usuario una interacción flexible con el chatbot.

Eje de usabilidad: eficiencia.

Justificación: la reducción del esfuerzo mental y conversacional del usuario es esencial, pues una buena comunicación permite concluir tareas en menor tiempo.

- H3 **Naturalidad:** se refiere a si el usuario puede percibir que está conversando con una computadora o bien con un humano.

Ejemplo: si el chatbot no logra entender al usuario, ofrece dirigirlo con una persona.

Beneficio: nunca confundir al usuario, pues se pueden ocasionar efectos adversos, como el del valle inquietante [43].

Eje de usabilidad: satisfacción.

Justificación: al disminuir la complejidad de las conversaciones aumenta el placer de usar el chatbot.

- H4 **Aprendizaje:** denota la capacidad que tiene el chatbot de aprender nuevas entradas e interacciones, así como de ofrecer alternativas cuando no comprendió al usuario.

Ejemplo: si el usuario solicitó un formulario de inscripción, pero el chatbot no entendió, le ofrece contactar al encargado de servicios escolares.

Beneficio: ya sea entrenamiento automático o manual, el aprendizaje será fundamental para que el chatbot se mantenga útil.

Eje de usabilidad: satisfacción.

Justificación: es indispensable incrementar constantemente la capacidad de aprendizaje del chatbot para que sea una herramienta dinámica y realmente útil.

- H5 **Funcionalidad:** expresa la facilidad de uso de cada una de las funciones que componen al chatbot.

Ejemplo: el chatbot ofrece un *widget* de calendario cuando el usuario ingresa una fecha.

Beneficio: evitar sobre o subdesarrollos en *widgets* o funciones que ya se han estudiado con anterioridad.

Eje de usabilidad: eficiencia.

Justificación: la reducción del esfuerzo mental y de los posibles errores de entrada son factores necesarios para lograr un flujo de trabajo natural.

E. Etapa de Validación

La validación se llevó a cabo con la ayuda de diez expertos en usabilidad (tres mujeres y siete hombres, de entre 30 y 50 años). Se escogió a los expertos por su experiencia realizando evaluaciones de usabilidad y UX. Todos los expertos son profesores universitarios con estudios de posgrado en alguna rama de la Computación; dos de ellos pertenecen a CINVESTAV-IPN. Su experiencia proviene tanto de la industria como de la academia (entre 5 y 25 años). Cabe resaltar que ninguno de ellos está directamente relacionado con esta investigación, con la obvia excepción de que fueron voluntarios para participar de este estudio.

Para esta etapa, se validó la propuesta contrastándola con el conjunto de heurísticas más popular, i.e., el propuesto por Nielsen y Molich [44] que, si bien fue ideado para la inspección de software, ha encontrado lugar en evaluaciones de diversos tipos de productos y servicios [24]. El conjunto original se componía de nueve heurísticas, no obstante, para

esta evaluación, se utilizó el conjunto nuevo que está conformado por 10 heurísticas [45]. Los expertos fueron divididos en dos grupos. Un grupo utilizó las heurísticas propuestas en este artículo, mientras que el otro usó las de Nielsen y Molich. Ambos grupos evaluaron un chatbot para la educación llamado *Ask Frank*. Este chatbot está implementado dentro de la plataforma Messenger de Facebook y responde a preguntas sencillas sobre Matemáticas, Ciencia e Historia [46].

La organización de cinco expertos por grupo se debe a que Nielsen recomienda integrar de tres a cinco expertos en una evaluación para obtener el máximo costo-beneficio [47].

La Fig. 1 muestra las asociaciones que hicieron los expertos de los problemas con las heurísticas. Como se puede observar, se considera que el grupo que trabajó con el conjunto de Nielsen y Molich acertó en un 100%, esto se debe tanto a la claridad de las heurísticas como a la familiarización que tienen los expertos con ellas. El grupo que evaluó utilizando las heurísticas para chatbots obtuvo, de manera similar, buenos resultados, pues se juzga que sus asociaciones problema-heurística son correctas en un 84.22% de los casos. Visto de otra manera, se sopesa que se equivocaron en un 15.78%, lo que indica que las heurísticas son entendibles.

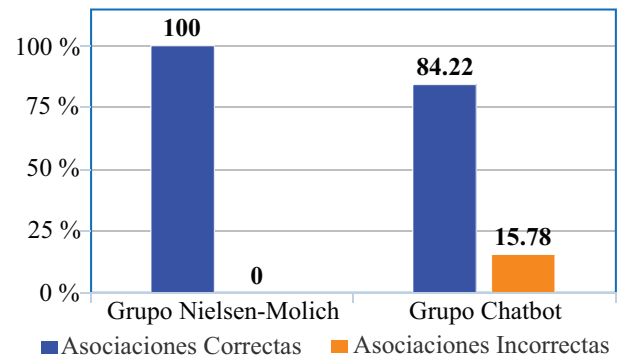


Fig. 1. Porcentajes de asociaciones correctas e incorrectas de acuerdo a cada grupo de heurísticas.

Otro indicio de que las heurísticas para chatbots son buenos indicadores para encontrar problemas, es el número de estos últimos que resultó de la evaluación del grupo que trabajó con dicha propuesta. Como se representa en la Fig. 2, los expertos que evaluaron con estas heurísticas pudieron encontrar 19 problemas, mientras que el grupo que evaluó con el conjunto de Nielsen y Molich encontró 8. Sólo hubo 4 problemas en la intersección.

F. Etapa de Refinación

Gracias a la retroalimentación que se obtuvo de los expertos, se decidió refinar la heurística H1 Completitud, pues podía llegarse a confundir con H2 Contexto. De esta manera, la heurística se modificó como se muestra a continuación:

- H1 **Completitud:** se refiere a la facultad del chatbot para obtener invariablemente todos los datos necesarios para resolver una tarea.

TABLA I
PROBLEMAS DE USABILIDAD Y SU SEVERIDAD

Problema	Heurísticas	Severidad
Tengo que poner la fecha completa y en un formato predeterminado (DD/MM/AA) para crear un recordatorio, no hay flexibilidad.	H1, H3, H4	2.4
Hacen falta algunos <i>widjets</i> (e.g., previsualizaciones de archivos) ya que interactuar solamente mediante texto se vuelve algo monótono.	H4	1.2
Algunas tareas requieren demasiados intentos para que el chatbot entienda lo que el usuario quiso decir.	H2, H3, H4	4.0
No tengo la confianza de proporcionarle mis datos personales.	H3, H4	2.8
Asignar calificaciones, alumno por alumno, es una tarea engorrosa.	H5	3.2
Al chatbot le hace falta «personalidad», i.e., básicamente es una máquina contestadora y sus respuestas son muy formales.	H2, H3	1.6
Se deben ofrecer más iniciativas, e.g., si pregunto por una calificación, me debería ofrecer ver el resto de mis calificaciones.	H4, H5	2.6
Debería poder integrarse con otras aplicaciones (e.g., <i>iCal</i>).	H5	1.6
En ocasiones es difícil cambiar de contexto; cuando se está llevando a cabo una tarea, el chatbot pierde el contexto de la tarea anterior.	H1, H4	3.2
No hay persistencia, i.e., si quiero llegar a cierto punto de una tarea, no hay atajos, es como volver a tener siempre la misma conversación por completo.	H2, H3, H4	2.8

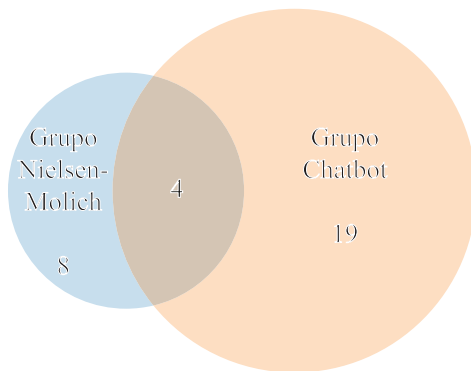


Fig. 2. Comparativa de problemas identificados por cada grupo de heurísticas.

V. CASO DE ESTUDIO

Una vez que las heurísticas propuestas fueron explicadas, así como sus orígenes y su validación, en esta sección se utilizan para evaluar un producto. La metodología que se siguió para esta evaluación fue la expuesta por Barnum [24] y replicada con éxito en diversos estudios [48], [49], [50].

Como objeto de evaluación se eligió un trabajo previo, un chatbot que sirve como herramienta extraescolar para alumnos de secundaria [51] (proyecto disponible en <https://github.com/jfuy/CinvestavCHATBOT>). Este chatbot es una herramienta que cuenta con diversas funcionalidades para maestros, alumnos y personal administrativo; los usuarios interactúan con él mediante texto, en español mexicano. El chatbot es capaz de almacenar y hacer recordatorios de fechas importantes (e.g., límite para envío de tareas, exámenes e inscripciones), ofrecer material a los alumnos (que los maestros previamente habían guardado), permitir el intercambio de archivos, responder preguntas frecuentes de procesos administrativos, entre otras funciones. El chatbot es una aplicación

web desarrollada con *AngularJS*, *Firebase*, *Dialogflow*, y *NodeJS*.

Para esta evaluación, se contó con la ayuda de los mismos cinco expertos que trabajaron con las heurísticas para chatbots en la Etapa de Validación (ver Sección IV-E), pues se quería aprovechar la experiencia que adquirieron en dicha etapa. En primer lugar, se les permitió familiarizarse un poco con el chatbot, su funcionamiento e interfaz de usuario. Después, se les pidió que realizaran algunas tareas sencillas (e.g., hacerle alguna pregunta al chatbot, agendar un examen y preguntar por una calificación) para después hacer una evaluación rápida, i.e., que dijeran cuáles eran los posibles errores que habían encontrado de acuerdo a las heurísticas que se proponen. Esto último sirvió para afianzar el significado de las heurísticas y, de esta manera, obtener mejores resultados. Para finalizar la sesión, la instrucción a los expertos fue que tenían dos días para usar el chatbot, detectar problemas y hacer una lista de las heurísticas violadas.

Una vez que todos los expertos hubieron evaluado el chatbot y elaborado sus listas, estas fueron consolidadas en una lista maestra de problemas. Esta lista maestra fue entregada a cada experto para que, individualmente, asignara una calificación, según la severidad del problema, y notara cuáles eran las heurísticas violadas en cada caso. Así, las calificaciones fueron promediadas y los resultados se presentan en la Tabla I. Las calificaciones se asignaron de acuerdo a la escala mostrada en la Tabla II [24].

A. Resultados

Con la ayuda de nuestras heurísticas, los expertos encontraron un total de 16 problemas (un promedio de 3.2 problemas por evaluador). La severidad de los problemas tuvo una media de 2.54. La lista maestra se compone de 10 problemas únicos, con un total de 20 violaciones a las heurísticas. Con 7 problemas encontrados, **H4 - Aprendizaje**

TABLA II
ESCALA DE SEVERIDAD PARA PROBLEMAS [24]

Calificación	Descripción
0	No es un problema.
1	Problema cosmético. No es necesario solucionarlo a menos de que se cuente con tiempo extra.
2	Problema menor. Tiene prioridad baja para solucionarse.
3	Problema mayor. Importante de solucionar, se le debe dar alta prioridad.
4	Problema catastrófico. Es imprescindible solucionarlo antes de lanzar el producto.

fue la heurística con mayor número de violaciones, seguida por **H3 - Naturalidad** con 5 violaciones. En contraste, con sólo 2 violaciones, **H1 - Completitud** fue la heurística con menor número de problemas detectados (ver Fig. 3).

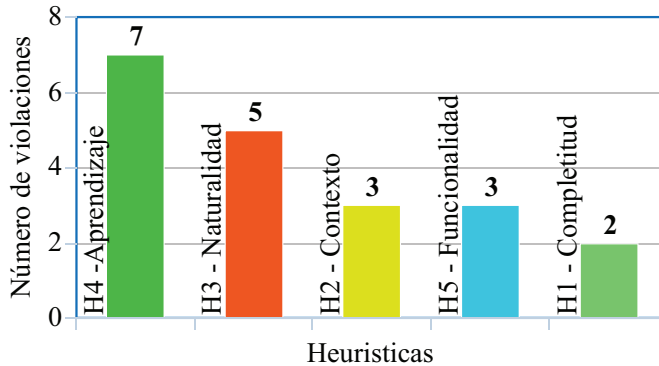


Fig. 3. Violaciones a las heurísticas para chatbots.

Con respecto a la severidad de los problemas detectados, se puede observar en la Fig. 4 que el nivel de severidad 3 - “Problema mayor” fue el más frecuente con un 42%. Al contrario, se nota que 0 - “No es un problema” obtuvo un 0%.

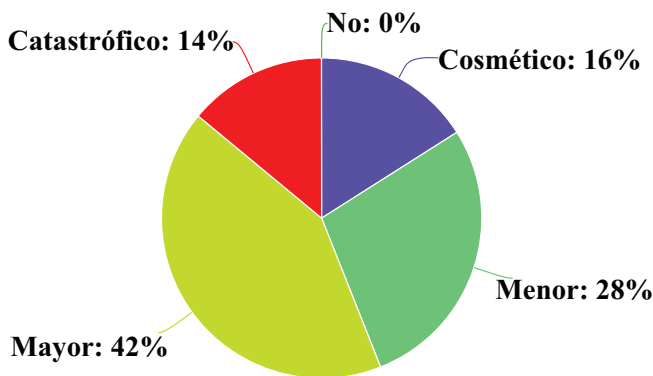


Fig. 4. Severidad de las violaciones.

B. Discusión

No es coincidencia que la heurística con mayor número de violaciones fuera **H4 - Aprendizaje**, pues se relaciona

directamente con la posibilidad de completar una tarea; si al chatbot le tomó demasiados intentos para comprender la entrada del usuario, este protocolo se vuelve desgastante, pues el proceso para dar comienzo una tarea se vuelve más costoso en todo sentido, que la tarea *per se*. Posiblemente, una manera de mitigar esto, además de mejorar el entrenamiento del chatbot, es ofrecer ligas que lleven a empezar eventos, i.e., que el chatbot tenga la iniciativa de comenzar tareas que se relacionen a la entrada del usuario.

Es interesante observar que **H2 - Contexto** y **H5 - Funcionalidad** obtuvieran el mismo número de violaciones. Por comentarios de los evaluadores, se puede deducir que este empate de calificaciones pueda tener cierta correlación; cuando los usuarios no saben si el chatbot sigue en el mismo contexto que ellos o que las respuestas brindadas no proveen una retroalimentación adecuada, crea una desconfianza de seguir con la tarea, especialmente si se trata con información privada. Finalmente, que **H1 - Completitud** haya sido la menos violada, indica que cuando el chatbot ayudó exitosamente a los usuarios con sus tareas, lo hizo porque sabía qué datos necesitaba para ello. El chatbot parece ser una herramienta indicada para el escenario de ayuda escolar.

En lo concerniente a la severidad de las violaciones, se puede argumentar que los evaluadores encontraron problemas que definitivamente deben ser corregidos, pues la mayor parte de las calificaciones, i.e., el 42%, cayó dentro de la clasificación de “Problema mayor”, de manera que es importante solucionarlos para mejorar la usabilidad del sistema. Por lo que se puede observar, gran parte de estos problemas se relaciona con los flujos de comunicación, i.e., al ser el texto la única forma de interacción que ofrece el chatbot, limita muchas tareas que se tienen que hacer de manera constante (e.g., asignar calificaciones a los alumnos). Como algunos expertos ya lo apuntaban, esto se podría solucionar con la implementación de *widgets* que ofrezcan al usuario un atajo hacia la tarea deseada, sin la necesidad de tener que pasar por la misma conversación una y otra vez.

Si bien, la clasificación de “Problema catastrófico” obtuvo un 14% de las calificaciones totales, se puede considerar un porcentaje relativamente alto. Los problemas detectados por los expertos revelan que aquellos que entran en este nivel de severidad tienen que ver con el entrenamiento e intentos del chatbot, i.e., con aspectos que conciernen al procesamiento y comprensión de lenguaje natural. La única solución previsible es mejorar el entrenamiento del chatbot.

El resto de los problemas obtuvo un 44%, i.e., la combinación de problemas “menores” y “cosméticos”. Puede que esto sea el resultado de los problemas más grandes, pues si los flujos de interacción fallan en cierto punto, lo más probable es que esto ocasione más fallas. Si bien es cierto que algunos se solucionarían con la implementación de ciertas características que los expertos echaron de menos, se sospecha que la mayoría de estos problemas se resolverán una vez que se lidie con los de mayor magnitud.

En general, se puede decir que la mayor parte de las funcionalidades del chatbot necesitan una mejoría considerable.

Una desventaja inherente a las evaluaciones heurísticas es el sesgo cognitivo [52]. Esto quiere decir que la calidad de las

evaluaciones está supeditada a la experiencia de los expertos; los problemas que encuentren están condicionados a lo que ellos ya tienen prejugado como «buena» usabilidad. Es por ello que, como ya se mencionó en la Sección I, una evaluación heurística es parte de un proceso integral para mejorar la usabilidad de un servicio o producto. No obstante, se tomaron medidas para tratar de mitigar los sesgos en las pruebas:

- Se adoptó una metodología estándar de evaluación.
- Se contó con la ayuda de cinco expertos, todos con amplia experiencia en el campo de la usabilidad y UX.
- Todos los expertos tuvieron la misma información como punto de partida.

Algunos indicadores cualitativos de que los resultados son válidos son la cantidad y las características de los problemas encontrados. Por una parte, hubo problemas repetidos entre expertos (es por ello que, de 16 problemas en total, se terminó con una lista maestra de tan sólo 10) lo que indica que entendieron las heurísticas y las aplicaron de manera similar. Por otra parte, los problemas son exactos y bien identificados, concurren con el propósito de las heurísticas propuestas, i.e., no se trata de razonamientos generales o vagos que pudieran encajar coincidentemente.

VI. CONCLUSIONES Y TRABAJO FUTURO

En este artículo se presentó un novel conjunto de heurísticas de usabilidad para evaluar chatbots. Mediante un caso de estudio que incluyó un chatbot para la educación y la ayuda de cinco expertos, se pusieron a prueba dichas heurísticas.

El caso de estudio reveló los problemas ocasionados por violar las heurísticas. Se estima que esta evaluación fue exitosa, pues se identificaron problemas concisos y relacionados directamente con las heurísticas.

Evaluar la usabilidad de un chatbot es un reto complejo. Una parte de la solución reside, en gran medida, en mejorar los mecanismos de inteligencia artificial de este tipo de sistemas, pero sería un enfoque simplista tomar sólo eso en cuenta. Otra parte importante es el contexto de uso del chatbot; lo más probable es que un mecanismo general de evaluación no baste para abarcar todos los contextos en los que un chatbot puede emplearse, pues mientras que en un escenario una característica es deseable, en otro puede ser todo lo contrario. No obstante, es evidente que algunos mecanismos siempre se pueden mantener independientes del contexto y es ahí en donde las heurísticas propuestas pueden resultar más útiles. Si los elementos base de interacción de un chatbot tienen una usabilidad positiva, esto puede llevar a menos problemas de UX en el futuro, cuando el chatbot adquiera características propias de su entorno.

Los expertos estuvieron satisfechos con las heurísticas, ya que les permitieron enfocarse en su evaluación e identificar problemas de manera más sencilla. Sin embargo, también sugirieron que hay que refinar las heurísticas, pues de este modo se pueden entender y aplicar de manera más efectiva. Esto representa un desafío importante en el futuro, pues tal como se pudo apreciar en la Sección IV-E, las heurísticas fueron malinterpretadas en un 15.78% de los casos (ver Fig. 1). Aunque se hizo un pequeño refinamiento a la heurística H1

(ver Sección IV-F), es necesario mejorarlas para que en toda ocasión resulten claras para todos los evaluadores.

Este trabajo puede ser considerado como una investigación de naturaleza exploratoria; se está al tanto de que una evaluación heurística no constituye una evaluación de usabilidad completa, sino que simplemente es una etapa de todo un proceso de evaluación y diseño. Se sabe que un solo estudio es insuficiente para obtener conclusiones sólidas, sin embargo, los resultados que se obtuvieron fueron satisfactorios, pues demuestran que, a pesar de que las heurísticas necesitan mejorar, es posible usarlas en una evaluación.

Como trabajo futuro, para enriquecer las heurísticas, se piensa hacer pruebas para obtener expectativas e ideales que las poblaciones más jóvenes tienen respecto a este tipo de sistemas (e.g., generaciones «Z» y «alfa»). De esta manera, no sólo se obtendrán puntos de vista de los grupos que probablemente más usan chatbots, sino también de usuarios no expertos.

Se estableció un primer paso para crear métodos de evaluación de usabilidad en el terreno poco explorado de los chatbots. Se juzga como prioridad identificar las posibles características que tienen mayor peso en la usabilidad de los chatbots, tanto las dependientes, como las independientes de contexto.

AGRADECIMIENTOS

El trabajo descrito en este artículo fue realizado gracias al financiamiento del «Fondo SEP-CINVESTAV de Apoyo a la Investigación (Convocatoria 2018)». Número de proyecto 120, titulado «Desarrollo de un chatbot inteligente para asistir el proceso de enseñanza/aprendizaje en temas educativos y tecnológicos».

REFERENCIAS

- [1] B. A. Shawar and E. Atwell, "Chatbots: are they really useful?" in *Ldv forum*, vol. 22, no. 1, 2007, pp. 29–49.
- [2] I. Intelligence, "Chatbot market in 2021: Stats, trends, and companies in the growing ai chatbot industry," Feb 2021. [Online]. Available: <https://www.businessinsider.com/chatbot-market-stats-trends>
- [3] R. Ravi, "Intelligent chatbot for easy web-analytics insights," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Sep. 2018, pp. 2193–2195.
- [4] A. V. Dian Sano, T. Daud Imanuel, M. Intanadias Calista, H. Nindito, and A. Raharto Condrobimo, "The application of agnes algorithm to optimize knowledge base for tourism chatbot," in *2018 International Conference on Information Management and Technology (ICIMTech)*, Sep. 2018, pp. 65–68.
- [5] B. R. Ranoliya, N. Raghuvanshi, and S. Singh, "Chatbot for university related faqs," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Sep. 2017, pp. 1525–1530.
- [6] H. Agus Santoso, N. Anisa Sri Winarsih, E. Mulyanto, G. Wilujeng saraswati, S. Enggar Sukmana, S. Rustad, M. Syaifur Rohman, A. Nugraha, and F. Firdausillah, "Dinus intelligent assistance (dina) chatbot for university admission services," in *2018 International Seminar on Application for Technology of Information and Communication*, Sep. 2018, pp. 417–423.
- [7] A. Argal, S. Gupta, A. Modi, P. Pandey, S. Shim, and C. Choo, "Intelligent travel chatbot for predictive recommendation in echo platform," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan 2018, pp. 176–183.
- [8] R. De', N. Pandey, and A. Pal, "Impact of digital surge during covid-19 pandemic: A viewpoint on research and practice," *International Journal of Information Management*, p. 102171, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0268401220309622>

- [9] A. S. Miner, L. Laranjo, and A. B. Kocaballi, "Chatbots in the fight against the covid-19 pandemic," *npj Digital Medicine*, vol. 3, no. 1, p. 65, May 2020. [Online]. Available: <https://doi.org/10.1038/s41746-020-0280-0>
- [10] K. Hao, "The pandemic is emptying call centers. ai chatbots are swooping in," May 2020. [Online]. Available: <https://www.technologyreview.com/2020/05/14/1001716/ai-chatbots-take-call-center-jobs-during-coronavirus-pandemic/>
- [11] P. Vergadia, "How can chatbots help during global pandemic (covid-19)?" Apr 2020. [Online]. Available: <https://medium.com/google-cloud/how-can-chatbots-help-during-global-pandemic-covid-19-9c1a4428d8c2>
- [12] C. J. Luo, V. Y. L. Wong, and D. E. Gonda, "Code free chatbot development: An easy way to jumpstart your chatbot!" in *Proceedings of the Seventh ACM Conference on Learning @ Scale*, ser. L@S '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 233–235. [Online]. Available: <https://doi.org/10.1145/3386527.3405932>
- [13] J. Ferreira, J. Noble, and R. Biddle, "Agile development iterations and ui design," in *Agile 2007 (AGILE 2007)*, 2007, pp. 50–58.
- [14] D. Gandasari, "Evaluation of online learning with digital communication media during the covid 19 pandemic," *Journal of the Social Sciences*, vol. 48, no. 3, 2020.
- [15] C. M. Barnum, "1 - establishing the essentials," in *Usability Testing Essentials*, C. M. Barnum, Ed. Boston: Morgan Kaufmann, 2011, pp. 9 – 23. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/B9780123750921000015>
- [16] H. Petrie and N. Bevan, "The evaluation of accessibility, usability, and user experience." *The universal access handbook*, vol. 1, pp. 1–16, 2009.
- [17] E. Law, V. Roto, A. P. Vermeeren, J. Kort, and M. Hassenzahl, "Towards a shared definition of user experience," in *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '08. Florence, Italy: Association for Computing Machinery, 2008, p. 2395–2398. [Online]. Available: <https://doi.org/10.1145/1358628.1358693>
- [18] E. L.-C. Law, V. Roto, M. Hassenzahl, A. P. Vermeeren, and J. Kort, "Understanding, scoping and defining user experience: A survey approach," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '09. Boston, MA, USA: Association for Computing Machinery, 2009, p. 719–728. [Online]. Available: <https://doi.org/10.1145/1518701.1518813>
- [19] P. Morville, "Experience design unplugged," in *ACM SIGGRAPH 2005 Web Program*, ser. SIGGRAPH '05. Los Angeles, California: Association for Computing Machinery, 2005, p. 10–es. [Online]. Available: <https://doi.org/10.1145/1187335.1187347>
- [20] O. Parlangei, E. Marchigiani, and S. Bagnara, "Multimedia systems in distance education: effects of usability on learning," *Interacting with Computers*, vol. 12, no. 1, pp. 37 – 49, 1999. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S095354389800054X>
- [21] J. Sandars, "The importance of usability testing to allow e-learning to reach its potential for medical education," *Education for Primary Care*, vol. 21, no. 1, pp. 6–8, 2010. [Online]. Available: <https://doi.org/10.1080/14739879.2010.11493869>
- [22] L. A. Ensina, H. D. Lee, W. S. R. Takaki, N. A. R. Maciejewski, N. Spolaôr, and F. C. Wu, "Heuristics-based responsiveness evaluation of a telemedicine computational web system," *IEEE Latin America Transactions*, vol. 17, no. 03, pp. 444–452, 2019.
- [23] A. R. Svaigen and L. A. F. Martimiano, "Netanimations mobile app: Improvement of accessibility and usability to computer network learning animations," *IEEE Latin America Transactions*, vol. 16, no. 1, pp. 272–278, 2018.
- [24] C. M. Barnum, "3 - big u and little u usability," in *Usability Testing Essentials*, C. M. Barnum, Ed. Boston: Morgan Kaufmann, 2011, pp. 53 – 81. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/B9780123750921000039>
- [25] R. H. R. Harper, "The role of hci in the age of ai," *International Journal of Human-Computer Interaction*, vol. 35, no. 15, pp. 1331–1344, 2019. [Online]. Available: <https://doi.org/10.1080/10447318.2019.1631527>
- [26] M. Ienca and E. Vayena, "On the responsible use of digital data to tackle the covid-19 pandemic," *Nature Medicine*, vol. 26, no. 4, pp. 463–464, Apr 2020. [Online]. Available: <https://doi.org/10.1038/s41591-020-0832-5>
- [27] S. Valtolina, B. R. Barricelli, and S. D. Gaetano, "Communicability of traditional interfaces vs chatbots in healthcare and smart home domains," *Behaviour & Information Technology*, vol. 39, no. 1, pp. 108–132, 2020. [Online]. Available: <https://doi.org/10.1080/0144929X.2019.1637025>
- [28] A. B. Kocaballi, L. Laranjo, and E. Coiera, "Understanding and Measuring User Experience in Conversational Interfaces," *Interacting with Computers*, vol. 31, no. 2, pp. 192–207, 05 2019. [Online]. Available: <https://doi.org/10.1093/iwc/iwz015>
- [29] R. Ren, J. W. Castro, S. T. Acuña, and J. de Lara, "Evaluation techniques for chatbot usability: A systematic mapping study," *International Journal of Software Engineering and Knowledge Engineering*, vol. 29, no. 11n12, pp. 1673–1702, 2019. [Online]. Available: <https://doi.org/10.1142/S0218194019400163>
- [30] H. Ding, N. Ranade, and A. Cata, "Boundary of content ecology: Chatbots, user experience, heuristics, and pedagogy," in *Proceedings of the 37th ACM International Conference on the Design of Communication*, ser. SIGDOC '19. Portland, Oregon: Association for Computing Machinery, 2019. [Online]. Available: <https://doi.org/10.1145/3328020.3353931>
- [31] S. Holmes, A. Moorhead, R. Bond, H. Zheng, V. Coates, and M. Mctear, *Usability Testing of a Healthcare Chatbot: Can We Use Conventional Methods to Assess Conversational User Interfaces?* New York, NY, USA: Association for Computing Machinery, 2019, p. 207–214. [Online]. Available: <https://doi.org/10.1145/3335082.3335094>
- [32] J. Yin, T.-T. Goh, B. Yang, and Y. Xiaobin, "Conversation technology with micro-learning: The impact of chatbot-based learning on students' learning motivation and performance," *Journal of Educational Computing Research*, 2020. [Online]. Available: <https://doi.org/10.1177/0735633120952067>
- [33] M. S. Rafael, T. B. L. María, F. U. Antonio, and D. L. F. M. Hanns, "Support to the learning of the chilean tax system using artificial intelligence through a chatbot," in *2019 38th International Conference of the Chilean Computer Science Society (SCCC)*, 2019, pp. 1–8.
- [34] D. Quiñones, C. Rusu, S. Roncagliolo, V. Rusu, and C. A. Collazos, "Developing usability heuristics: A formal or informal process?" *IEEE Latin America Transactions*, vol. 14, no. 7, pp. 3400–3409, 2016.
- [35] J. Weizenbaum, "Eliza—a computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, p. 36–45, Jan. 1966. [Online]. Available: <https://doi.org/10.1145/365153.365168>
- [36] J. Sedoc, D. Ippolito, A. Kirubarajan, J. Thirani, L. Ungar, and C. Callison-Burch, "ChatEval: A tool for chatbot evaluation," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 60–65. [Online]. Available: <https://www.aclweb.org/anthology/N19-4011>
- [37] K. Kuksenok and A. Martyniv, "Evaluation and improvement of chatbot text classification data quality using plausible negative examples," 2019.
- [38] C. Segura, À. Palau, J. Luque, M. R. Costa-Jussà, and R. E. Banchs, "Chatbot, a chatbot for the spanish "la liga,"" in *9th International Workshop on Spoken Dialogue System Technology*, L. F. D'Haro, R. E. Banchs, and H. Li, Eds. Singapore: Springer Singapore, 2019, pp. 319–330.
- [39] M. Qiu, F.-L. Li, S. Wang, X. Gao, Y. Chen, W. Zhao, H. Chen, J. Huang, and W. Chu, "AliMe chat: A sequence to sequence and rerank based chatbot engine," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 498–503. [Online]. Available: <https://www.aclweb.org/anthology/P17-2079>
- [40] H. Kazi, B. Chowdhry, and Z. Memon, "Medchatbot: An umls based chatbot for medical students," *International Journal of Computer Applications*, vol. 55, no. 17, 2012.
- [41] G. C. Guerino and N. M. C. Valentim, "Usability and user experience evaluation of conversational systems: A systematic mapping study," in *Proceedings of the 34th Brazilian Symposium on Software Engineering*, ser. SBES '20. Natal, Brazil: Association for Computing Machinery, 2020, p. 427–436. [Online]. Available: <https://doi.org/10.1145/3422392.3422421>
- [42] N. Bevan, J. Carter, J. Earthy, T. Geis, and S. Harker, "New iso standards for usability, usability reports and usability measures," in *Proceedings, Part I, of the 18th International Conference on Human-Computer Interaction. Theory, Design, Development and Practice - Volume 9731*. Berlin, Heidelberg: Springer-Verlag, 2016, p. 268–278. [Online]. Available: https://doi.org/10.1007/978-3-319-39510-4_25
- [43] L. Ciechanowski, A. Przegalinska, M. Magnuski, and P. Gloor, "In the shades of the uncanny valley: An experimental study of human-chatbot interaction," *Future Generation Computer Systems*, vol. 92, pp. 539 – 548, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167739X17312268>
- [44] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '90. Seattle, Washington, USA:

Association for Computing Machinery, 1990, p. 249–256. [Online]. Available: <https://doi.org/10.1145/97243.97281>

- [45] J. Nielsen, “10 usability heuristics for user interface design,” Nov 2020. [Online]. Available: <https://www.nngroup.com/articles/ten-usability-heuristics/>
- [46] P. Smutny and P. Schreiberova, “Chatbots for learning: A review of educational chatbots for the facebook messenger,” *Computers & Education*, vol. 151, p. 103862, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360131520300622>
- [47] J. Nielsen, “Finding usability problems through heuristic evaluation,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '92. Monterey, California, USA: Association for Computing Machinery, 1992, p. 373–380. [Online]. Available: <https://doi.org/10.1145/142750.142834>
- [48] J. Zhang, T. R. Johnson, V. L. Patel, D. L. Paige, and T. Kubose, “Using usability heuristics to evaluate patient safety of medical devices,” *Journal of Biomedical Informatics*, vol. 36, no. 1, pp. 23 – 30, 2003.
- [49] N. K. Chuan, A. Sivaji, and W. F. W. Ahmad, “Usability heuristics for heuristic evaluation of gestural interaction in hci,” in *Design, User Experience, and Usability: Design Discourse*, A. Marcus, Ed. Cham: Springer International Publishing, 2015, pp. 138–148.
- [50] B. A. Kumar, M. S. Goundar, and S. S. Chand, “A framework for heuristic evaluation of mobile learning applications,” *Education and Information Technologies*, Jan 2020. [Online]. Available: <https://doi.org/10.1007/s10639-020-10112-8>
- [51] S. Mendoza, M. Hernández-León, L. M. Sánchez-Adame, J. Rodríguez, D. Decouchant, and A. Meneses-Viveros, “Supporting student-teacher interaction through a chatbot,” in *Learning and Collaboration Technologies. Human and Technology Ecosystems*, P. Zaphiris and A. Ioannou, Eds. Cham: Springer International Publishing, 2020, pp. 93–107.
- [52] U. G. S. Administration, “Heuristic evaluations and expert reviews,” Oct 2013. [Online]. Available: <https://www.usability.gov/how-to-and-tools/methods/heuristic-evaluation.html>



José Rodríguez obtained in 2005 a PhD degree in Computer Science at the Université Paul Sabatier Toulouse III, France. Since 2006 he has been a full-time researcher at the Department of Computer Science of CINVESTAV-IPN, Mexico City, Mexico. His lines of research are Distributed Systems and Ubiquitous Computing.



Amilcar Meneses-Viveros obtained in 2009 a PhD degree in Computer Science at CINVESTAV-IPN, Mexico City, Mexico. Since 2010 he works as a full-time researcher in that department. His lines of research include Human-Computer Interaction and Parallel Computing.



Luis Martín Sánchez-Adame obtained in 2013 a BS in Computer Engineering at Durango Institute of Technology, Durango, Mexico. And in 2016 a MSc degree at the Department of Computer Science of CINVESTAV-IPN (Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional), Mexico City, Mexico. Currently, he is a PhD student at the aforementioned department. His lines of research include Human-Computer Interaction, User Interface Design, User Experience, and Social Media.



Sonia Mendoza obtained a PhD degree in Computer Science at INPG (Institut National Polytechnique de Grenoble), Grenoble, France. She is a full professor at the Department of Computer Science at CINVESTAV-IPN, Mexico City, Mexico. She has been published in several international journals and conferences, and has advised around 31 MSc theses and five PhD dissertations. Her major research interests are in Computer-Supported Cooperative Work, Ubiquitous Computing, and Human-Computer Interaction.



José Fidel Urquiza-Yllescas is a research tutor professor of the computer class at IEMS, Vasco de Quiroga Campus, Mexico City. Previously, he worked for ten years as a professor of the Faculty of Sciences of UNAM teaching the computer class for the Physics major. He studied Computer Science at UNAM. Subsequently, he obtained a master's degree at UAM-I. Currently, he is a PhD student in the Department of Computer Science at CINVESTAV-IPN, Mexico City. His main research areas are Architecture and Design of Chatterbots and Human-

Computer Interaction.