Journal of Computer Science

Review

# Usable AI: Critical Review of Its Current Issues and Trends

**Majed Alshamari and Thanaa Alsalem**

*Information Systems Department, King Faisal University, Hofuf, Saudi Arabia*

*Article history*
Received: 29-09-2022
Revised: 21-11-2022
Accepted: 28-01-2023

Corresponding Author:
Majed Alshamari
Information Systems Department, King Faisal University, Hofuf, Saudi Arabia
Email: smajed@kfu.edu.sa

**Abstract:** Usable Artificial Intelligence (AI) application refers to an AI solution that is characterized by the easiness of use and learning through an optimum interface established by a proficient Human-Computer Interface (HCI) design. This article reviews the related literature to find out the importance of HCI to make AI valuable and usable. The article also investigates the possible solutions and guidelines that ensure the usability of AI-powered applications as well as the corresponding challenges of designing usable AI. It also essentially explores several usability evaluation methods employed by recent studies. The findings demonstrate the substantial role of HCI in designing Human-centred AI (HAI) applications. HAI lets AI applications support humans instead of replacing them and grants user's better control over security and privacy. No standard usability measures exist and more research is required especially on the issues of interpretability, integrability, and collaboration.

**Keywords:** Usable Artificial Intelligence (AI), Human Computer Interaction (HCI), Human-Centred AI (HAI), Usability

## Introduction

Artificial Intelligence (AI) and Human-Computer Interface (HCI) are converging terms that refer to computing and intelligent behavior. AI strategies have caught the interest of a growing number of HCI researchers and machine learning applications are increasingly discussed in the HCI-related literature. Highly developed AI applications require contributions from the HCI community to define and improve the interaction between AI and users to support the creativity aspect of humans instead of replacing it. Usability is considered a system quality attribute. A usable AI application is identified as an AI solution that is easy to use and learn through an optimum interface established by a proficient HCI design (Xu, 2019; Schmidt *et al*., 2021). User Interface (UI) allows users to control and interact with applications in a user-friendly interface, which impact the User experience (UX) (Akinsola *et al*., 2021; Kristiadi *et al*., 2017). The prevalence of AI in different sectors entails putting humans at the center of AI developing lifecycles and designing Human-centred AI (HAI) applications (Bond *et al*., 2019). Designing HAI should follow guidelines and be tested through various usability evaluation methods that address the usability of intelligent applications rather than non-intelligent ones. The AI-based applications are endless and more and more research is required to study and develop standard usability evaluation methods. Consequently, this research aims to explore this issue theoretically based on reviewing the related literature.

## The Significance of HCI Design in AI Applications

HCI is concerned with the human aspect of the interaction that occurs among individuals and AI-based applications instead of algorithmic performance. It focuses on creating a positive influence on human users and society (Schmidt *et al*., 2021). UI and UX are two terms related to HCI. UI allows users to control and interact with applications in a user-friendly interface (Akinsola *et al*., 2021). It focuses on the style and looks, representing the mutual point where interaction occurs between users and the system design and functionalities. UI involves three components: Voice-controlled User Interfaces (VUI), Graphical User Interfaces (GUI), and Gesture-centred Interfaces. As a consequence, companies that develop web applications and mobile apps give UI a high priority to enhance the overall user experience (Kristiadi *et al*., 2017). On the other hand, UX depicts and centers on the user's experience when an interaction happens with the system. Interaction tends to be broader than an interface as it allows communication and dialog between the user and the system (Akinsola *et al*., 2021). UX involves three components: The user, the application, and interactions.

Since HCI specialists are keen on developing user-cantered designs, they are required to apply this user-centredness to AI solutions (Bond *et al*., 2019). AI-based applications are increasingly prevailing in personal coaching, healthcare, and education, to name a few. This prevalence entails putting humans at the center of AI-developing lifecycles, referring to the so-called HAI (Bond *et al*., 2019). An interactive HAI "is an artificial intelligence that enables interactive exploration and manipulation in real-time and is designed with a clear purpose for human benefit while being transparent about who has control over data and algorithms" (Schmidt, 2020). HAI reshapes AI by changing AI systems and algorithms with humans at the center as depicted in Fig. 1 (Shneiderman, 2020).

Building HAI applications brings many advantages. HAI allows users to think, act, create, and see in exceptional ways, by integrating robust user experiences with AI methods to assist in services that people need (Shneiderman, 2020). HAI ensures that AI-based applications augment and enhance humans instead of replacing them. Therefore, an HAI approach overcomes the issues of out-of-control technologies, calms fears of robot-driven unemployment, and gives users better control over security and privacy (Shneiderman, 2020). It also brings opportunities in terms of ethics, adoption, usability, and avoidance of the unintended detrimental consequences of AI systems (Bond *et al*., 2019). The significance of HAI can be highlighted in the next incidents. For instance, a recent UX test, based on voice interactions, was conducted for three intelligent assistants of leading brands in the US market. The test demonstrated success only in some simple tasks and failure in all sorts of complicated tasks. Moreover, recently, autonomous vehicles have also been involved in several deadly accidents attributed partially to HCI design issues. These incidents emphasize the significant role of HCI design in building usable AI (Xu, 2019).

To conclude, HCI designers have a significant role in introducing usable AI applications and developing HAI solutions that ensure the human aspect of the interaction between AI and users.

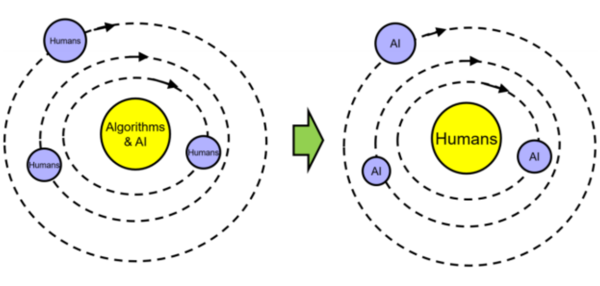


**Fig. 1:** HAI reshapes AI (Shneiderman, 2020)

## Usability Solutions and Guidelines

Building usable AI requires developing and following a set of guidelines (Amershi *et al*., 2019). In intelligent applications, the prototyping of the UI should focus on interactivity instead of visuals as used in non-intelligent applications. Designers need to consider AI first, develop more intuitive UIs, design a dynamic human-machine functional allocation, and rank the use of intelligent functions, such as real-time smart search, behavior, voice input, and contextual information, to decrease repetitive human activities (Xu, 2019). Therefore, many researchers attempted to introduce dedicated guidelines that guide the design of intelligent systems. For instance, Amershi created a thorough set of guidelines (Amershi *et al*., 2019) in which they collected eighteen commonly applicable design guidelines. Those guidelines included the following: "Make clear what the system can do, make clear how well the system can do what it can do, time services based on context, show contextually relevant information, match relevant social norms, mitigate social biases, support efficient invocation, support efficient dismissal, support efficient correction, scope services when in doubt, make clear why the system did what it did, remember recent interactions, learn from user behavior, update and adapt cautiously, encourage granular feedback, convey the consequences of user actions, provide global controls and notify users about changes" (Amershi *et al*., 2019). Another example is that of (Funk *et al*., 2020) who added three main guidelines for overcoming response delays in automotive user interfaces, depending on empirical data. These guidelines included "do not interrupt users while making voice input, always respond within the sweet-spot response range and inform users about possible response delays" (Funk *et al*., 2020).

Usable AI applications are critical. Currently, AI applications are utilized in the healthcare industry for various usage of patient treatments, and even more for surgical procedures. Having usable AI applications can minimize risks in such critical AI applications. Another example is how AI is used in the financial industry. It can maximize the detection rate of different malicious activities in the banking sector. AI also can clearly contribute to investigating the application of machine learning to medicine and report the diagnostic performance and caution of machine learning in radiology (Sharma *et al*., 2022).

Sharma *et al*. (2022) contributed effectively to the area of intelligent sensing as it discusses different dimensions such as projects and their areas. It also involves different algorithms. These are critical to be highlighted in intelligent sensing.

The previously mentioned guidelines are helpful in designing HAI. Following these guidelines helps develop responsible and productive AI-based applications that are qualified to be actors in society. To ensure the usability of HAI applications, various usability evaluation tests can be carried out as discussed in the next section.

## Usability Evaluation Methods of AI Applications

Balsa *et al*. (2020) evaluated the usability of an intelligent virtual assistant application. For testing the usability experimentally, eleven patients and nine nurses were purposely selected. For data collection, the System Usability Scale (SUS) was applied and augmented with digital summaries, open questions, follow-ups, and diaries through telephone. The patients were asked to try a list of prescribed tasks for 26 days at home. Those tasks included, for instance, talking to the virtual assistant daily. The nurses were required to evaluate the application individually for 8-10 days and interact with the application on the same day.

Harms (2019) introduced a virtual reality usability evaluation method that can be applied rapidly and completely automated with no need of asking participants to perform pre-specified tasks. Rather, it records the real performance of a virtual reality application and generates task trees. Then, these task trees are analyzed automatically to find usability issues, such as user behavior. The researchers conducted a big case study and the findings demonstrated the capability of the automated approach in reporting detailed usability issues. The introduced approach is capable of handling the various interactions of modalities, for instance, controller, hand interaction, or gaze. Also, it can determine users' inefficiency; however, it cannot identify the misunderstandings of users.

To *et al*. (2021) evaluated the usability, feasibility, and effectiveness of chatbots powered with a Natural Language Processor (NLP) to engage and interact with users to monitor their physical activity by receiving data from a smartphone application. Without a control group, a quasi-experimental design was performed and data were collected at baseline and after six weeks. As part of the self-monitoring process, each participant wore an activity tracker sensor that linked to the chatbot through a smartphone app. The chatbot informs each participant of their physical activity level and adjusts the daily goals accordingly. It also sends out motivational messages daily and provides information about the physical activity's benefits. The chatbot checks the step counts/min that were completed daily as well. Usability and acceptability information was self-recorded. The main test criteria were the number of step counts provided by the activity tracker sensor and participants.

Zwakman *et al*. (2021) evaluated the usability of AI-enabled voice assistants, relying on a user-based usability evaluation. They believed that SUS is one of the well-known usability instruments in the GUI area, but it cannot cover the differences between GUI and systems based on voice. They accordingly questioned the suitability of SUS for testing voice-based systems. To test the usability evaluation of voice assistants, they devised and applied a subjective scale, called the Voice Usability Scale (VUS), along with SUS to test the distinguished aspects of a voice-based application. By collecting data from 62 participants, the exploratory Factor Analysis confirmed several pitfalls of SUS for measuring the usability of an AI-enabled voice assistant. On the other hand, VUS recognized three aspects, namely, effectiveness, visibility, and recognisability.

Di Nuovo *et al*. (2019) conducted a preliminary usability evaluation of two cognitive systems powered by AI. The systems are the IBM cloud AI "Watson" and the social robot "Pepper", both being multimodal interfaces. The test involved 62 participants who used the system, free of any assistance, and completed the SUS questionnaire. The findings showed that both systems are highly reliable in terms of usability.

Maza-Jimenez and Torres-Carrión (2020) evaluated the usability of an AI-powered conversational system based on efficiency, effectiveness, and satisfaction. A virtual assistant called "max" the virtual assistant of the private technical university of Loja was used as a case study. The participants were chosen to represent the university community including students (15), parents (15), and outsiders (15). The evaluation was conducted virtually. The effectiveness, efficiency, and satisfaction metrics were determined based on the International Standards for Organization (ISO) 9241-11. To collect data, an initial test was conducted as an ad-hoc survey to gather personal-related data and determine the participant's emotional state. This initial test included personal data, informed consent, emotions classification, the cause of the emotions, and the effectiveness and efficiency of the virtual assistant. The second instrument was used to evaluate the general satisfaction of users through the system usability questionnaire, the Spanish version of the Computer Systems Usability Questionnaire (CSUQ), which consists of 16 components. The findings ensured low levels of effectiveness, efficiency, and satisfaction, and thus the interaction strategy of Max needed further checking.

Holzinger *et al*. (2020) believed that it is essential to ensure explain the ability to build effective and efficient AI applications that have interactive human-AI interfaces. Accordingly, they tried to test to what extent an AI application is explainable. They introduced a System Causability Scale (SCS) for measuring the explanations' quality. They applied the concept of CA usability introduced by Holzinger *et al*. (2019) and combined it with SUS.

Guo *et al*. (2020) investigated the usability of a conversational AI agent in smart homes. The usability metrics included task completion time, perceived system usability, and the number of queries utilized in completing tasks. Different levels of task complexity (high vs. low) and different types of conversation (absence and presence of unceasing conversation) were tested. Data were collected from 18 participants. The findings showed differences in the metrics with different complexity levels and different conversation types. Complex tasks required extra time for task completion and a greater number of queries per task, calling for more usable conversational agents.

Maguire *et al*. (2019) evaluated speech interfaces through a set of heuristics built on the well-known general heuristics developed by Nielsen and Molich for user interface design. From a design perspective, those heuristics were found to be extremely rooted in a GUI-based environment.

Lane *et al*. (2019) introduced a usability testing model. The authors tested and compared the usability of two question-answering systems based on an NLP. The systems' modalities included text and speech. The proposed model utilized and integrated two frameworks, namely, SUS and the System Usability Metric (SUM). The model also considered context-specific determination. To identify the weight of each model's metric, key informant interviews were conducted.

Bogers *et al*. (2019) evaluated the usability of an Intelligent Personal Assistant (IPA), namely, Siri. Based on a purposive sampling method, 20 participants were recruited, including both experienced and inexperienced users with IPAs in general. Each participant performed a user-based summative usability testing that was prolonged for 25 min. The usability tests were recorded in audio and video files for an afterward analysis. First, a pilot test was performed on two participants. Then, the basic test included seven different tasks with one simple task as a pre-test training. After completing the tasks, the SUS questionnaire was used in addition to some questions related to Siri's performance as well as the exerted effort during the test.

Langevin *et al*. (2021) proclaimed that there is an absence of an established set of usability heuristics that evaluate and guide the conversational agents' design. Accordingly, they introduced a number of heuristics based on Nielsen's heuristics plus expert feedback. Their test was performed on a voice-based personal assistant and a chatbot. The findings indicated that the proposed heuristics were more effective than Nielsen's heuristics and helped participants to identify more usability issues, especially concerning interaction design, human-like characteristics, dialogue content, help and guidance, and data privacy.

Dutsinma *et al*. (2022) identified the usability measures currently utilized for voice assistants through a systematic review. They identified a comprehensive list of usability measures, which included the measuring metrics of the ISO 9241-11 framework and additional outside measures. The authors confirmed that numerous aspects have not yet been studied. Additionally, the current ISO 9241-11 framework is not appropriate for measuring the latest technological expansion of AI owing to the user expectation and needs that have changed with the continuous advancement of technology. The measures of the ISO 9241-11 framework cannot explain certain usability measures, for instance, cognitive load, machine voice, and attitude.

Pal *et al*. (2019) evaluated the usability of Siri and Alexa through a mixed methodology approach comprising real-world testing and an online questionnaire. The data were collected from 275 questionnaires and 52 field tests, using two groups: Native and non-native speakers. The questionnaire was designed to ask the participants about their regular inquiries, such as asking about the weather and/or directions, checking e-mail/messages, playing music, monitoring other smart home equipment, or placing orders for items. The highly ranked activities were chosen for the usability test. Then, an extended SUS questionnaire was distributed to test the user experience. It was found that voice assistants were usable for both groups. However, the native English user was more satisfied. The voice assistant understood non-English words spoken in English with difficulty. It was also incapable of distinguishing between numerous voices.

Rivero Jiménez *et al*. (2021) evaluated the usability of a voice assistant by using three tools. SUS was used because it is commonly used and free. To make the testing process more consistent, the researchers also used two other scales validated by the portuguese institute of electronics and telematics engineering of Aveiro: The ICF-US I Scale and the ICF-US II scale (Martins *et al*., 2016). The ICF-US I scale identifies the general usability problems. The ICF-US II scale identifies the possible facilitators and/or barriers and locates any element that requires further work to enhance the device.

Gates *et al*. (2019) performed a usability test for three machine learning applications developed to help in the screening stage of titles and abstracts in systematic reviews. The test was done for two different scenarios: A complete automated simulation for excluding irrelevant results and a semi-automated simulation for supporting a single reviewer's work. User experiences were evaluated for each tool. The researchers conducted three systematic reviews based on two retrospective screening simulations. The three applications were Abstrackr, DistillerSR, and Robot Analyst. A set of 200 records was used for training and identifying the relevance of the whole set of records. The percentage of missed studies, workload, and time savings were set against two separate human-performed screening procedures. For testing user experience, eight research staff members applied the tools and filled in a SUS survey.

Dahri *et al*. (2019) sought to evaluate the usability of an AI-based mobile health application. Fifteen patients participated to perform various tasks. The efficiency (time spent), success rate, errors, and Satisfaction (SUS) were recorded. In addition, the ISO 9241-11 typical metrics were employed. The findings showed the capability of the utilized tests in identifying usability issues that can be avoided during the design of the application. Notably, both the mobile know-how and educational level influenced the usability scores attributed to the consumed time, occurrence of task errors, and completion of tasks.

Table 1 demonstrates a summary of the reviewed studies in terms of the author (s), targeted AI system, and usability evaluation method.

**Table 1:** Literature review summary

| Study | System | Method |
|---|---|---|
| Balsa *et al*. (2020) | Intelligent virtual assistant | Extended SUS, digital notes, diaries, and follow-ups through telephone |
| Harms (2019) | A virtual reality application | Recording real performance and generating task trees are analyzed automatically to find usability issues |
| To *et al*. (2021) | Chatbot | Self-reporting usability test |
| Zwakman *et al*. (2021) | Voice assistants | VUS, SUS |
| Di Nuovo *et al*. (2019) | IBM cloud AI "Watson" and the social robot "Pepper" | SUS |
| Maza-Jimenez and Torres-Carrión (2020) | Virtual assistant (Max) | ISO 9241-11, CSUQ |
| Holzinger *et al*. (2020) | Human-AI systems | SCS, SUS |
| Guo *et al*. (2020) | Conversational AI agent in a smart home context | Paper-based questionnaire to gather responses to perceived system usability measures |
| Maguire *et al*. (2019) | Speech interfaces | Heuristics based on Nielsen's heuristics |
| Lane *et al*. (2019) | Question Answering (QA) systems | SUS, SUM |
| Bogers *et al*. (2019) | Voice assistant (Siri) | Usability test, extended SUS |
| Langevin *et al*. (2021) | Conversational agents | Heuristics based on Nielsen's heuristics and expert feedback |
| Dutsinma *et al*. (2022) | Voice assistant | ISO 9241-11 framework and additional outside measures |
| Pal *et al*. (2019) | Voice assistants (Siri and Alexa) | Questionnaire, usability test, SUS |
| Rivero Jiménez *et al*. (2021) | Voice assistant | SUS, ICF-US I, ICF-US II (Bogers *et al*., 2019) |
| Gates *et al*. (2019) | Three machine learning applications | Extended SUS |
| Dahri *et al*. (2019) | AI-based mobile health application | ISO 9241-11, SUS |

In short, it is noticed that most researchers rely on SUS, ISO 9241-11, and/or Nielsen's heuristic framework to test how usable an AI-based application is. They typically extend these frameworks with additional questions that suit the nature of AI technology. Noticeably, not many researchers evaluated the usability of AI applications automatically.

## Usability Challenges of AI Applications

In the field of HCI, various usability evaluation methods were originally developed for all applications. This brings several usability challenges when it comes to testing the usability of intelligent applications. The first challenge is presentation and interaction. Choosing an appropriate interaction style is challenging as the interface is composed of conversational interaction and humanlike virtual agents plus conventional GUI solutions. Such systems need to consider any change in the initial dialogue and the possibility of interruption of user activities (Ziegler, 2019). It is essential to consider the addition of the learning capabilities of AI-based applications and how human-computer interaction involves the meaning of teaming and integration (Xu, 2019). Additionally, AI cooperates with users and thus AI applications are expected to support users ultimately but not purely substitute them. This cooperation is essential, for instance, in mutual learning in collaborative work (e.g., robotics) and handover processes (e.g., automated driving) (Ziegler, 2019). HCI designers should take into consideration that humans and machines are collaborative partners and teammates. Having two cognitive agents with the machine's enhanced capability as it learns over time makes the HCI design of AI applications more complex (Xu, 2019). Another challenge is the lack of transparency and explanation ability of automated decisions. To have a usable AI-based application, developing systems for explaining the decision from the user's perspective is essential. An additional challenge is the controllability of the decision-making processes. Usable applications should allow user control to start at various stages: During the selection of data, selection of algorithms, and even the feedback. Usability entails a closer interaction between the user and algorithm and more across-the-board intervention possibilities. Finally, keeping ethical and legal aspects is a bedrock of usable AI applications. Since AI applications perform cognitive tasks, a variety of questions arise, among them the decisions that have no explanation. Having a biased or discriminatory decision that could happen intentionally or unintentionally is a real problem. For instance, in loan granting or application procedures, having a biased decision is ethically and legally critical (Ziegler, 2019).

Developing HAI is not easy and involves many challenges, including presentation and interaction, cooperation rather than substitution, transparency, and explaining the ability of automated decisions and controllability over the decision-making processes, as well as ethical and legal aspects.

## Future Work and Research Directions

Microsoft assures that designing interactive experiences in HAI is a significant area that entails further research to enhance communication between people and AI systems (Horvitz, 2019). As discussed earlier, HAI faces many challenges. Moreover, the review of the related literature demonstrated a remarkable absence of standard usability measures. Most studies relied on SUS, ISO 9241-11, or Nilsen's heuristics as a basis for testing HAI usability. They augmented these frameworks with additional related questions that revolve around the interaction of human-AI applications. However, AI applications are endless, and further usability research is essential for making such intelligent applications usable and valuable. In response, numerous researchers shared various ideas for future research. For instance, Dutsinma *et al*. (2022) argued that the current ISO 9241-11 framework is not appropriate for measuring the usability of AI-based applications. It is not capable of clearly explaining machine voice, cognitive load, and attitude. Therefore, they encouraged research on building usability frameworks based on ISO 9241-11 as a bedrock.

For advancing the field of HAI generally Inkpen *et al.* (2019) proposed a range of guiding questions. For instance, from the explain ability and interpretability perspective, researchers are required to find out "how can users explore AI systems' results and logic to identify non-obvious failure modes?" In terms of integrating AI and humans, a research question would be "what challenges might surface in attempting to do so?" From the collaborative decision-making perspective, a question can be "how do we ensure that when there is a human-in-the-loop such as in complex or life-changing decision-making they remain critical and meaningful while creating and maintaining an enjoyable user experience?" In the area of the HCI design process for AI, we can ask "how can algorithmic tools be made more readily accessible to those whose expertise lies outside of machine learning?"

Furthermore, incorporating ethical principles into the development process is needed to be further investigated. AI applications are expected to adopt and preserve different ethical principles such as data governance, fairness, transparency, and accountability. Such principles should be adopted as early as possible and closely monitored during the development process.

## Conclusion

The prevalence of AI-based applications brings the issue of AI usability. HCI allows AI applications to be used for humans and HAI puts humans at the center of the AI development lifecycle. In this way, AI applications support humans rather than replace them. However, HAI applications should follow a set of guidelines to produce useful applications. For instance, they should focus on interactivity instead of visuals as used in non-intelligent applications. AI applications should necessarily undergo a usability evaluation test to ensure that all the specified guidelines are applied. SUS, ISO 9241-11, and/or Nielsen's heuristic framework are basically used to test how usable an AI-based application is. These frameworks are typically extended to cover the additional questions that AI applications arouse. Although some researchers resort to automatically testing the usability, it is not a ubiquitous method. In fact, designing HAI involves many challenges, including presentation, interaction, cooperation, transparency, controllability, ethics, and law. The intersection of HCI and AI is an affluent area where researchers and practitioners can contribute to a wide range of related topics. Developing appropriate and standard usability measures, explaining ability and interpretability, integrability, collaboration and the HCI design process for AI are all examples of possible areas that require further research.

## Acknowledgment

Thank you to the publisher for their support in the publication of this research article. We are grateful for the resources and platform provided by the publisher, which have enabled us to share our findings with a wider audience. We appreciate the efforts of the editorial team in reviewing and editing our work, and we are thankful for the opportunity to contribute to the field of research through this publication.

## Funding Information

The authors have not received any financial support or funding to report.

## Author's Contributions

**Majed Alshamari:** Formalized the idea, critically review the references, and drafted, revision finalized the paper.

**Thanaa Alsalem:** Critically reviewed the published papers, wrote the paper, and formatted the final review of the paper.

## Ethics

I undersigned that this article has not been published elsewhere. The authors declare no conflict of interest.

## References

Akinsola, J. E. T., Akinseinde, S., Kalesanwo, O., Adeagbo, M., Oladapo, K., Awoseyi, A., ... & Heimgartner, R. (2021). *Application of artificial intelligence in user interfaces design for cyber security threat modeling* (pp. 1-28). Intech Open. ISBN-10: 9781839689666.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1-13). https://doi.org/10.1145/3290605.3300233

Balsa, J., Félix, I., Cláudio, A. P., Carmo, M. B., Silva, I. C. E., Guerreiro, A., ... & Guerreiro, M. P. (2020). Usability of an intelligent virtual assistant for promoting behavior change and self-care in older people with type 2 diabetes. *Journal of Medical Systems*, *44*, 1-12. https://doi.org/10.1007/s10916-020-01583-w

Bogers, T., Al-Basri, A. A. A., Ostermann Rytlig, C., Bak Møller, M. E., Juhl Rasmussen, M., Bates Michelsen, N. K., & Gerling Jørgensen, S. (2019). A study of usage and usability of intelligent personal assistants in Denmark. In *Information in Contemporary Society: 14th International Conference, iConference 2019, Washington, DC, USA, March 31–April 3, 2019, Proceedings 14* (pp. 79-90). Springer International Publishing. https://doi.org/10.1007/978-3-030-15742-5_7

Bond, R. R., Mulvenna, M. D., Wan, H., Finlay, D. D., Wong, A., Koene, A., ... & Adel, T. (2019, October). Human Centered Artificial Intelligence: Weaving UX into Algorithmic Decision Making. In *RoCHI* (pp. 2-9).

Dahri, A. S., Al-Athwari, A., & Hussain, A. (2019). Usability evaluation of mobile health application from AI perspective in rural areas of Pakistan. https://doi.org/10.3991/ijim.v13i11.11513

Di Nuovo, A., Varrasi, S., Conti, D., Bamsforth, J., Lucas, A., Soranzo, A., & McNamara, J. (2019, March). Usability evaluation of a robotic system for cognitive testing. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 588-589). IEEE. https://doi.org/10.1109/hri.2019.8673187.

Dutsinma, F. L. I., Pal, D., Funilkul, S., & Chan, J. H. (2022). A Systematic Review of Voice Assistant Usability: An ISO 9241–11 Approach. *SN Computer Science*, *3*(4), 267. https://doi.org/10.1007/s42979-022-01172-3.

Funk, M., Cunningham, C., Kanver, D., Saikalis, C., & Pansare, R. (2020, September). Usable and Acceptable Response Delays of Conversational Agents in Automotive User Interfaces. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 262-269). https://doi.org/10.1145/3409120.3410651

Gates, A., Guitard, S., Pillay, J., Elliott, S. A., Dyson, M. P., Newton, A. S., & Hartling, L. (2019). Performance and usability of machine learning for screening in systematic reviews: A comparative evaluation of three tools. *Systematic Reviews*, *8*(1), 1-11. https://doi.org/10.1186/s13643-019-1222-2.

Guo, J., Tao, D., & Yang, C. (2020). The effects of continuous conversation and task complexity on usability of an AI-based conversational agent in smart home environments. In *Man–Machine–Environment System Engineering: Proceedings of the 19th International Conference on MMESE* (pp. 695-703). Springer Singapore. https://doi.org/10.1007/978-981-13-8779-1_79

Harms, P. (2019). Automated usability evaluation of virtual reality applications. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *26*(3), 1-36. https://doi.org/10.1145/3301423

Holzinger, A., Carrington, A., & Müller, H. (2020). Measuring the quality of explanations: The system causability scale (SCS) comparing human and machine explanations. *KI-Künstliche Intelligenz*, *34*(2), 193-198. https://doi.org/10.1007/s13218-020-00636-z.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(4), e1312. https://doi.org/10.1002/widm.1312

Horvitz, E. (2019). A shared vision to advance Human-Centered AI. Microsoft Research. https://doi.org/10.1145/3290607.3299002

Inkpen, K., Chancellor, S., De Choudhury, M., Veale, M., & Baumer, E. P. (2019, May). Where is the human? Bridging the gap between AI and HCI. In *Extended abstracts of the 2019 chi conference on human factors in computing systems* (pp. 1-9).

Kristiadi, D. P., Udjaja, Y., Supangat, B., Prameswara, R. Y., Warnars, H. L. H. S., Heryadi, Y., & Kusakunniran, W. (2017, November). The effect of UI, UX and GX on video games. In *2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)* (pp. 158-163). IEEE. https://doi.org/10.1109/CYBERNETICSCOM.2017.8311702

Lane, D., Renwick, R., McAvoy, J., & O'Reilly, P. (2019). The Advent of Speech Based NLP QA Systems: A Refined Usability Testing Model. In *Design, User Experience and Usability. Practice and Case Studies: 8th International Conference, DUXU 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part IV 21* (pp. 152-163). Springer International Publishing. https://doi.org/10.1007/978-3-030-23535-2_11

Langevin, R., Lordon, R. J., Avrahami, T., Cowan, B. R., Hirsch, T., & Hsieh, G. (2021, May). Heuristic evaluation of conversational agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-15). https://doi.org/10.1145/3411764.3445312

Maguire, M. (2019). Development of a heuristic evaluation tool for voice user interfaces. In *Design, User Experience and Usability. Practice and Case Studies: 8th International Conference, DUXU 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part IV 21* (pp. 212-225). Springer International Publishing. https://doi.org/10.1007/978-3-030-23535-2_16

Martins, A. I., Queirós, A., Silva, A. G., & Rocha, N. P. (2016, December). ICF based Usability Scale: Evaluating usability according to the evaluators' perspective about the users' performance. In *Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion* (pp. 378-383). https://doi.org/10.1145/3019943.3019997

Maza-Jimenez, L., & Torres-Carrión, P. (2020). Usability evaluation of a Conversational Artificial Intelligence Interface. https://ceur-ws.org/Vol-2747/paper11.pdf

Pal, D., Arpnikanondt, C., Funilkul, S., & Varadarajan, V. (2019, July). User experience with smart voice assistants: The accent perspective. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE. https://doi.org/10.1109/icccnt45670.2019.8944754

Rivero Jiménez, B., Conde Caballero, D., Jesús-Azabal, J., Luengo-Polo, J., Bonilla-Bermejo, J., & Mariano Juárez, L. (2021). Qualitative Research in Evaluation. An Usability Evaluation Protocol for the Assistant on Care and Health Offline (ACHO). In *Gerontechnology III: Contributions to the Third International Workshop on Gerontechnology, IWoG 2020, October 5-6, 2020, Évora, Portugal* (pp. 43-53). Springer International Publishing. https://doi.org/10.1007/978-3-030-72567-9_5

Schmidt, A. (2020, September). Interactive human centered artificial intelligence: A definition and research challenges. In *Proceedings of the International Conference on Advanced Visual Interfaces* (pp. 1-4). https://doi.org/10.1145/3399715.3400873

Schmidt, A., Giannotti, F., Mackay, W., Shneiderman, B., & Väänänen, K. (2021, August). Artificial intelligence for humankind: A panel on how to create truly interactive and Human-Centered AI for the benefit of individuals and Society. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part V* (pp. 335-339). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-85607-6_32

Sharma, A., Sharma, V., Jaiswal, M., Wang, H. C., Jayakody, D. N. K., Basnayaka, C. M. W., & Muthanna, A. (2022). Recent Trends in AI-Based Intelligent Sensing. *Electronics*, *11*(10), 1661. https://doi.org/10.3390/electronics1110166

Shneiderman, B. (2020). Human-centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction*, *12*(3), 109-124. https://doi.org/10.17705/1thci.00131

To, Q. G., Green, C., & Vandelanotte, C. (2021). Feasibility, Usability and Effectiveness of a Machine Learning–Based Physical Activity Chatbot: Quasi-Experimental Study. *JMIR mHealth and uHealth*, *9*(11), e28577. https://doi.org/10.2196/28577

Xu, W. (2019). Toward human-centered AI: a perspective from human-computer interaction. *interactions*, *26*(4), 42-46. https://dl.acm.org/doi/fullHtml/10.1145/3328485

Ziegler, J. (2019, June). Challenges in User-Centered Engineering of AI-based Interactive Systems. In *EICS Workshops* (pp. 51-55). https://ceur-ws.org/Vol-2503/paper1_8.pdf

Zwakman, D. S., Pal, D., & Arpnikanondt, C. (2021). Usability evaluation of artificial intelligence-based voice assistants: The case of Amazon Alexa. *SN Computer Science*, *2*, 1-16. https://doi.org/10.1007/s42979-020-00424-4