Journal of SAINTEKS
Vol 4, No 1, November 2026, page 81-91
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

Combining Takaful Domain and UI/UX Expertise: A Dual-Expert Review of a Shariah-Compliant Conversational Agent for Islamic Financial Literacy among Marginalized Communities

Azizah Che Omar*

School of Multimedia Technology & Communication (SMMTC), College of Art and Sciences, Universiti Utara Malaysia, Kedah, Malaysia
Email: *co.azizah@uum.edu.my
Corresponding Author Email: co.azizah@uum.edu.my

Abstract−*Shariah*-compliant chatbots are increasingly used to explain Islamic financial products such as *Takaful*, yet their interfaces are typically modelled on generic chat applications and evaluated only after release through end-user questionnaires. This practice is problematic for marginalized users with uneven financial, digital and language literacy, who are the least able to articulate why an interface fails, and in a faith-based domain a failure of clarity is also a failure of trust. Usability evaluation alone cannot judge whether content, tone and branding suit the *Shariah* context, while domain review alone cannot diagnose interaction problems. To address this gap, this study proposes and applies a dual-expert review that combines *Takaful* domain experts from the Malaysian Takaful Association (MTA) with academic UI/UX experts in the formative evaluation of SCAI, a bilingual mobile conversational agent. Domain experts were engaged through a focus group after a prototype demonstration, while UI/UX experts tested the prototype independently and were interviewed with an eleven-item protocol mapped to QUIS, USE and Chatbot Usability Scale (BUS) constructs. Thematic and root-cause analysis yielded four themes: visual design and branding, navigation and interface interaction, language consistency and accessibility, and conversational interaction. Each theme was traced to implemented refinements through a construct-mapped traceability matrix and triangulated against end-user ratings of the refined application (n = 33). The two expert groups surfaced complementary rather than redundant issues: domain experts judged legitimacy, whereas UI/UX experts judged operability. The study contributes four design principles for *Shariah*-sensitive conversational interfaces, namely legitimacy cues, provenance transparency, linguistic completeness and expectation setting, and argues that in this domain usability is part of how users judge *Shariah* assurance rather than a cosmetic concern.

Keywords: Dual-Expert Review; Shariah-Compliant Chatbot; Conversational Agent; Islamic Financial Literacy; Takaful; Formative Usability Evaluation; Chatbot Usability Scale

1. INTRODUCTION

Islamic financial literacy in Malaysia remains uneven, particularly among young Muslims, lower-income B40 households and marginalized Muslim communities whose adoption of digital services is itself limited (Osman et al., 2024; Munisamy et al., 2022; Nawang et al., 2025; Busari, 2025). Rather than revisit why such literacy matters, which is well established, this paper addresses a narrower and less examined problem: how the interface of a conversational agent should be designed so that these users can actually obtain trustworthy *Takaful* knowledge from it.

Designing for these users raises interface challenges that general-purpose chat applications do not resolve. Users with limited literacy and little experience of smartphone applications depend on visible controls, consistent labels and spoken interaction rather than dense text and hidden functions (Medhi et al., 2011). Users who move between *Bahasa Melayu* and English are disadvantaged when an interface is only partly localised. Conversational agents add a further difficulty: their capabilities are invisible until tried, so users form expectations from the first exchange and abandon agents whose scope is unclear (Luger & Sellen, 2016; Rapp et al., 2021). In a *Shariah*-sensitive domain these challenges are compounded by trust. Users must believe not only that an answer is correct but that it is *Shariah*-compliant, while large language models are known to produce fluent but unsupported statements (Ji et al., 2023; Kasneci et al., 2023). Because trust in chatbots depends on perceived expertise, transparency and the credibility of the organisation behind them (Nordheim et al., 2019; Følstad et al., 2021), interface elements such as branding, source attribution and language consistency carry institutional and religious weight, not only usability weight.

Current practice in the design of educational chatbots does not address these challenges well. Systematic reviews show that educational chatbots are evaluated mainly after development, through learner questionnaires, experiments and user-based testing (Okonkwo & Ade-Ibijola, 2021; Wollny et al., 2021; Kuhail et al., 2023; Kumar et al., 2024), with little reporting on how the interface was shaped before release. Established instruments such as QUIS and the USE Questionnaire were not designed to capture conversational qualities, which motivated the Chatbot Usability Scale (BUS) (Borsci et al., 2022), yet even BUS contains no construct for whether content and presentation are appropriate for a faith-based context. In the *Takaful* domain, research has examined why consumers adopt chatbots (Mohd Thas Thaker et al., 2026) but not how a *Takaful* chatbot should be designed. This status quo has two weaknesses. First, deferring evaluation to end users places the burden of discovery on those least able to articulate interface problems, since marginalized users are likely to experience an unexplained chatbot or a half-translated screen as personal failure rather than report it as a design flaw. Second, usability specialists cannot judge *Shariah* appropriateness and domain experts cannot diagnose interaction problems, so single-perspective reviews leave part of the design space unexamined.

The problem is urgent because *Takaful* operators and industry bodies are moving towards AI-based education and customer channels, and design errors in this domain are costly: an interface that obscures where an answer comes from,

Copyright © 2026 the author, Page 81
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4, No 1, November 2026, page 81-91
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

or mixes languages, can undermine confidence not only in the application but in the institution and the *Shariah* legitimacy it represents. Formative expert review addresses such problems at the stage where changes are cheapest (Elwy et al., 2020; Nielsen, 1994), but there is little published evidence on how domain and usability experts can be combined, how their findings differ and how their recommendations can be traced into design decisions for a *Shariah*-compliant chatbot.

This study was conducted within the development of SCAI, a *Shariah*-Compliant Conversational Agent built in collaboration with the Malaysian Takaful Association (MTA) to promote Islamic financial literacy and *Takaful* awareness among marginalized communities. SCAI is an Android application developed with React Native Expo that combines structured *Takaful* topic modules, an artificial intelligence chatbot supported by a curated knowledge base, bilingual *Bahasa Melayu* (BM) and English (EN) interaction, and voice input. A companion summative study reported high user acceptance of the final application across usefulness, ease of use, ease of learning, satisfaction, interface quality, accessibility and trust (Annamalai et al., 2026). The present paper reports the formative expert review that preceded that evaluation and produced the final design.

Accordingly, this study proposes and applies a dual-expert review to the initial SCAI prototype, combining *Takaful* content experts from MTA with academic UI/UX experts. It asks what issues each expert group identifies, why these issues arose, and how each recommendation can be traced to a design refinement and checked against end-user ratings. The contribution is threefold. Methodologically, it demonstrates a dual-expert formative review in which an interview protocol mapped to QUIS, USE and BUS constructs and a traceability matrix make expert feedback auditable. Empirically, it shows that domain and UI/UX experts surface complementary classes of problems in a *Shariah*-compliant chatbot. Theoretically, it derives four design principles for conversational interfaces in *Shariah*-sensitive domains, namely legitimacy cues, provenance transparency, linguistic completeness and expectation setting, which extend existing chatbot usability constructs to account for institutional and religious trust.

2. RESEARCH METHOD

2.1 Research Framework

This study adopted a qualitative formative evaluation design embedded within the design and development of the SCAI mobile prototype. The evaluation was conducted in Malaysia during the development phase, after a functional prototype had been completed and before the summative evaluation with end users. Its defining feature is a dual-expert review in which two kinds of expertise examine the same artefact through different lenses. The first source was a content expert group of 6 representatives of the Malaysian Takaful Association (MTA), the client and industry partner of the project, who judged whether the content, tone and branding were appropriate for the *Takaful* and *Shariah* context. The second source comprised two academic UI/UX experts with backgrounds in multimedia, interface design, user experience and human–computer interaction, who judged whether the interface could be found, understood and operated. The rationale for the size and composition of the panel, and the steps taken to strengthen the trustworthiness of the findings, are set out in Section 2.6.

Because the study is qualitative and formative, it was guided by evaluation questions rather than statistical hypotheses. The questions asked (i) how experts perceive the usability, visual design, navigation, accessibility and conversational interaction of the SCAI prototype, (ii) which issues each expert group identifies, why these issues arose and which refinements the experts recommend, and (iii) how these recommendations can be traced to design changes in the final prototype and whether the targeted areas were rated favourably by end users. The focal variables were the usability constructs adapted from QUIS (Chin et al., 1988), the USE Questionnaire (Lund, 2001) and BUS (Borsci et al., 2022), and the data were analysed using thematic analysis (Braun & Clarke, 2021). The research framework is shown in Figure 1.

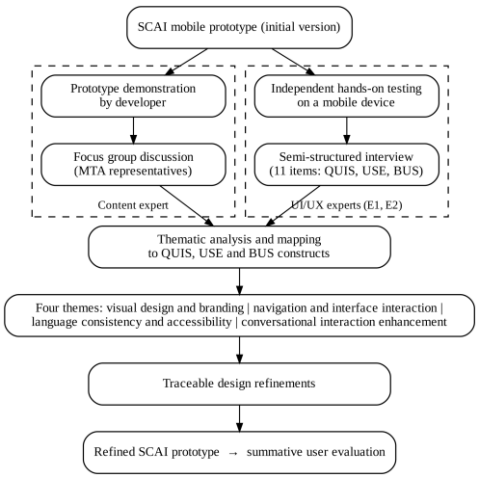


Figure 1. Research framework of the SCAI formative expert review

Copyright © 2026 the author, Page 82
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4, No 1, November 2026, page 81-91
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

2.2 Research Stages

The expert review was carried out in six stages. In the first stage, the experts were selected purposively against explicit inclusion criteria. MTA representatives were invited as content experts because of their domain expertise in *Takaful*, Islamic finance and *Shariah*-compliant educational content (each with at least 5 years of experience in the *Takaful* industry), while the two UI/UX experts were selected for their academic expertise in multimedia, interface design and usability evaluation (each with Phd holder and at least 5 ears of teaching or research in HCI). In the second stage, appointments were scheduled so that each expert had sufficient time to interact with the prototype and provide considered feedback. In the third stage, the developer demonstrated the prototype to the MTA representatives, explaining the application flow, the topic modules, the bilingual interaction and the chatbot functions, after which a focus group discussion (Krueger & Casey, 2015) was held on the educational presentation, *Shariah*-compliant communication, branding alignment and overall interface suitability.

In the fourth stage, each UI/UX expert explored the prototype independently on a mobile device, examining topic navigation, chatbot interaction, language switching, overlays, voice input and accessibility elements on the basis of their professional judgement. In the fifth stage, a semi-structured interview (DiCicco-Bloom & Crabtree, 2006) was conducted with each UI/UX expert after the hands-on session using the instrument described in Section 2.3. In the sixth stage, all feedback was documented, analysed thematically, mapped to the source constructs and translated into design refinements that were implemented in the final prototype before the summative evaluation; the refined areas were then compared with end-user ratings from that evaluation (Section 2.6).

2.3 Expert Interview Instrument

The interview instrument consisted of eleven open-ended questions adapted from three established instruments, as shown in Table 1. QUIS items addressed general impressions, the arrangement of information on screen, the sequence of screens, the ease of finding and selecting screen items, character shapes and the use of colour (Chin et al., 1988). The USE Questionnaire contributed items on ease of use and satisfaction (Lund, 2001), while BUS contributed items on the ease of starting a conversation and the visibility of and access to chatbot functions (Borsci et al., 2022). Each item was classified as Retained (R) when its wording remained close to the original construct or Adapted (A) when it was rephrased to refer to specific SCAI components such as the language overlay, topic cards and voice input. The open-ended format allowed experts to describe problems, explain their reasoning and propose concrete improvements.

Because the instrument is an open-ended interview protocol rather than a psychometric scale, internal-consistency indices such as Cronbach's alpha are not applicable to it. Its rigour rests instead on content validity, that is, whether each item adequately and unambiguously represents the construct to which it is mapped (Lynn, 1986; Polit & Beck, 2006). Three steps were taken. First, every item was mapped explicitly to a construct of a validated source instrument (Table 1), and items were classified as retained or adapted so that the degree of departure from the source is transparent.

Table 1. Adapted expert interview instrument

No.	Expert interview question	Source	Construct	Type
1	What is your overall feedback on the user interface design of the SCAI mobile prototype?	QUIS	Part E: General impressions	A
2	Is the opening flow (opening screen, language overlay, topic screen) clear and easy for first-time users to start using the application?	BUS	Ease of starting a conversation	A
3	How would you evaluate the arrangement of topic cards, overlays and the chatbot tab?	QUIS	Arrangement of information on screen	A
4	How effective are the buttons and tabs in supporting smooth navigation throughout the application?	QUIS	Screen items are easy to select	A
5	Are the buttons and icons such as send, voice input, close icon and language switch easy to identify?	QUIS	Items are easy to find	R
6	Do the font size, spacing and text layout support readability for users with different literacy levels?	QUIS	Character shapes (fonts)	A
7	How suitable are the colour usage, contrast and visual layout in making the interface aesthetically pleasing?	QUIS	Use of colours	A
8	Does the bilingual BM/EN interface switch improve ease of interaction for users?	USE	Ease of use	A
9	How effective is the chatbot screen layout in presenting the conversation history, prompt box and voice input?	BUS	Access to chatbot and visibility	A
10	Are there any interface components or screen flows that may confuse users during navigation?	QUIS	Sequence of screens	A
11	What specific improvements would you recommend for refining the user interface design before final development?	USE	Satisfaction	A

Note: R = Retained (wording close to the original construct); A = Adapted (wording modified to fit the SCAI interface).

Copyright © 2026 the author, Page 83
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4, No 1, November 2026, page 81-91
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

2.4 Expert Panel

The profile of the expert panel is summarised in Table 2. To preserve confidentiality, the MTA content expert group (n=6) is referred to as CE and the two UI/UX experts as E1 and E2. The content expert group evaluated the prototype after a guided demonstration, whereas the UI/UX experts evaluated it independently, which allowed the review to capture both an informed domain perspective and an unguided first-use perspective. This deliberate contrast between guided and unguided use also serves as a form of source triangulation (Section 2.6).

Table 2. Profile of the expert panel

Code	Expert source	Area of expertise	Method and evaluation focus
CE	Content expert group (MTA representatives)	*Takaful*, Islamic finance and *Shariah*-compliant educational content	Prototype demonstration and focus group discussion; educational presentation, *Shariah*-compliant communication, branding alignment, bilingual interaction
E1	UI/UX expert (academic)	Multimedia, user interface design and usability evaluation	Independent testing and semi-structured interview; interface usability, navigation flow, accessibility, visual consistency, conversational design
E2	UI/UX expert (academic)	User interface design, user experience and human–computer interaction	Independent testing and semi-structured interview; onboarding, interface familiarity, navigation discoverability, accessibility, bilingual consistency, chatbot usability

2.5 Data Analysis

Feedback from the focus group discussion and the interviews was documented and consolidated into a single feedback log. The log was analysed using thematic analysis through the phases of familiarisation, coding, theme development and theme review (Braun & Clarke, 2021). Statements that referred to the same interface element or problem were coded together, and related codes were grouped into candidate themes. The themes were then mapped deductively to the QUIS, USE and BUS constructs from which the interview items were derived, so that each theme could be interpreted against an established usability dimension. Visual design and navigation primarily reflected QUIS constructs on screen layout, colour and the sequence of screens; language consistency reflected the USE construct of ease of use; and conversational interaction reflected BUS constructs on starting a conversation and accessing chatbot functions. For each theme, the analysis then asked why the issue had arisen by tracing it to the design decision in the initial prototype that produced it, distinguishing issues inherited from generic chatbot conventions, issues caused by incomplete implementation and issues arising from domain requirements that had not been specified. Finally, a traceability matrix was constructed linking each theme, construct and recommendation to the refinement implemented in the prototype, and before-and-after screenshots were captured to document each change.

2.6 Panel Adequacy, Trustworthiness and Triangulation

The small panel was a deliberate choice for a formative review rather than an attempt at statistical generalisation, and its adequacy rests on three grounds. First, in expert inspection the yield of new problems diminishes quickly after the first few evaluators, and three to five evaluators typically uncover most interface problems (Nielsen & Landauer, 1993; Nielsen, 1994). Second, in qualitative research adequacy is judged by information power rather than headcount; a narrow aim, highly specific expertise and in-depth dialogue about a single artefact all increase the information each participant contributes (Malterud et al., 2016). Third, the content expert source was itself a group of 6 MTA representatives, so the review drew on the judgement of more than three individuals.

Nonetheless, because different evaluators detect different problems (Hertzum & Jacobsen, 2001), the findings were strengthened through three forms of triangulation (Carter et al., 2014). Method triangulation combined a focus group, hands-on inspection and individual interviews. Source triangulation contrasted domain with UI/UX expertise, and guided with unguided use. End-user triangulation compared each expert theme with the corresponding construct ratings and open-ended comments of 33 end users in the summative evaluation of the refined application (Annamalai et al., 2026), which is reported in Section 3.8. In addition, the traceability matrix and before-and-after screenshots form an audit trail that allows readers to verify how each interpretation led to a design decision (Lincoln & Guba, 1985).

3. RESULTS AND DISCUSSION

Overall, the experts perceived the initial SCAI prototype as clean, simple and intuitive. The arrangement of topic cards, the chatbot tab and the bottom navigation supported smooth movement through the application, the bilingual BM/EN switch was received positively as an accessibility feature, and the chat layout was considered familiar because it resembles widely used generative AI chatbots. Nevertheless, the thematic analysis yielded four themes of issues and recommendations: visual design and branding, navigation and interface interaction, language consistency and

Copyright © 2026 the author, Page 84
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4, No 1, November 2026, page 81-91
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

accessibility, and conversational interaction enhancement. Table 3 summarises the strengths and issues associated with each theme together with the construct to which it was mapped. For each theme, Sections 3.1 to 3.4 report what the experts observed, analyse why the issue arose in the initial prototype and describe the refinement implemented; Sections 3.6 to 3.8 then interpret the findings beyond the specific fixes.

Table 3. Summary of expert findings by theme

Theme	Mapped construct	Strengths noted	Issues and recommendations
Visual design and branding	QUIS: arrangement of information, use of colours	Clean and simple layout; clear topic cards	Slightly monochromatic and visually minimal; weak MTA identity; adopt MTA blue, add low-opacity branding and Islamic decorative elements, simplify header logo
Navigation and interface interaction	QUIS: sequence of screens, ease of selecting items	Simple, intuitive structure; smooth movement between topics and chatbot	Plain "X" close icon; small overlay container; crowded chat input area; no settings tab or direct MTA access
Language consistency and accessibility	USE: ease of use	Bilingual BM/EN interaction valued	Buttons, tabs and content not fully localised after switching; text-only voice transcription feedback; no font or display options
Conversational interaction enhancement	BUS: starting a conversation, access to chatbot functions	Familiar chat layout; clear prompt area and conversation flow	No introduction of chatbot capabilities; sources not shown; response language not always consistent; no external links, voice-to-voice or video support

3.1 Visual Design and Branding

The experts agreed that the initial interface was clean and uncluttered, but they described it as slightly monochromatic and visually minimal, with little to connect the application to MTA or to the Islamic context of its content (Figure 2a and 2b). Branding alignment had been one of the topics of the focus group with the content expert group, and it emerged as a shared concern across the panel because SCAI is intended to serve as an extension of MTA's public awareness efforts. The recommendations were to adopt MTA's corporate blue as the dominant colour, to add low-opacity MTA graphics and subtle Islamic decorative elements to the topic background, and to simplify the SCAI logo in the header. These recommendations were mapped to the QUIS constructs of arrangement of information on screen and use of colours.

The underlying cause lies in the visual convention the prototype inherited. Its neutral, near-monochrome style followed general-purpose generative AI chat applications, which are deliberately brand-agnostic because they serve every topic and every user. In that setting neutrality signals openness; in a *Shariah*-sensitive domain it removes the cue that tells users who stands behind an answer. Studies of online credibility show that users judge trustworthiness heavily from visual design and visible institutional affiliation (Fogg et al., 2003), and trust in chatbots likewise depends on the perceived credibility of the organisation behind them (Nordheim et al., 2019). The branding issue was therefore not a missing finishing touch but an unstated requirement: institutional and religious legitimacy had not been specified as a design objective, and the domain experts, who represent that institution, were best positioned to name it.

In response, the topic tab was redesigned with the MTA blue theme, a low-opacity decorative motif behind the topic cards and a simplified emblem-only logo in the header (Figure 2c). The topic cards were also given a tinted fill and more generous spacing, which strengthened the visual hierarchy between the header, the cards and the navigation bar. Beyond aesthetics, the change matters for credibility. In a domain where users must trust that information is accurate and *Shariah*-compliant, a visible association with a recognised industry body acts as a cue of legitimacy, which illustrates why the perspective of the client is needed alongside that of usability specialists.

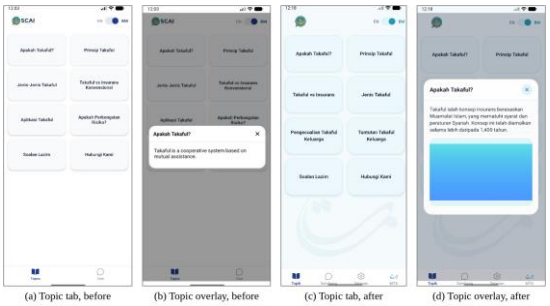


Figure 2. Topic Tab and Topic Overlay Before and After Refinement

3.2 Navigation and Interface Interaction

The navigation structure was generally rated as simple, intuitive and easy to understand. However, the UI/UX experts identified several interaction details that reduced comfort and clarity. The topic overlay used a small container and a plain

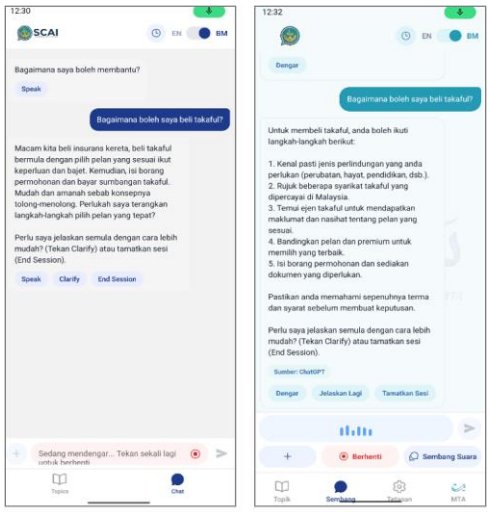
Copyright © 2026 the author, Page 85
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4, No 1, November 2026, page 81-91
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

"X" icon to close it (Figure 2b), which the experts considered insufficiently visible on a mobile screen. The chat input grouped several interaction functions within a single area, which made the input region feel crowded (Figure 3a). The experts also noted that the application lacked a place for user preferences and offered no direct route to MTA's own resources. These observations correspond to the QUIS constructs of the sequence of screens and the ease of finding and selecting screen items.

These issues share a common origin. As functions such as attachment, stop and voice conversation were added to the prototype, they were placed into existing containers rather than given their own space, so the interface became denser while each control became smaller. The two-tab navigation reflected a reasonable commitment to minimalism, but minimalism achieved by omitting functions hides them rather than simplifying them. For novice and lower-literacy smartphone users, visible, spatially separated and clearly labelled controls are more usable than compact multi-function areas, and such users are less likely to explore in search of hidden options (Medhi et al., 2011). The UI/UX experts, who tested the prototype without guidance, encountered these problems directly, whereas the content experts, who were shown the flow by the developer, were less exposed to them; this difference illustrates why unguided inspection was a necessary part of the review.

Four refinements were made. The overlay container was enlarged and the close icon was replaced with a circular, button-shaped control that is easier to see and tap (Figure 2d). The chat input was reorganised into a two-row layout, with the message field on the first row and the attachment, stop and voice conversation actions on the second (Figure 3b). The bottom navigation bar was expanded from two tabs to four by adding a Settings tab and a direct link to the MTA website alongside the Topics and Chat tabs (Figure 4). Expanding the navigation bar slightly increases the number of choices on screen, but each tab corresponds to a distinct user goal, which keeps the structure flat and predictable.



(a) Chatbot tab, before (b) Chatbot tab, after

Figure 3. Chatbot tab before and after refinement



Figure 4. Bottom navigation bar before and after refinement

3.3 Language Consistency and Accessibility

The bilingual BM/EN interaction was positively received during the evaluation sessions, and the experts regarded it as important for reaching users with different language preferences. The review nevertheless revealed that localisation was incomplete. When BM was selected, several interface elements remained in English, including the navigation labels "Topics" and "Chat", the overlay text and the chatbot action buttons "Speak", "Clarify" and "End Session" (Figures 2b, 3a and 4a). Such mixing weakens the very accessibility benefit that the bilingual feature is meant to provide, particularly for users who rely on a single language. The experts also pointed out that during voice input the only feedback was a line of text, which made the transcription process difficult to follow, and they recommended accessibility options for text size and display mode. These issues were mapped to the USE construct of ease of use.

The pattern of what was and was not translated reveals the cause. Topic content and chatbot responses switched language, whereas navigation labels, overlay text and action buttons did not, which indicates that localisation had been

Copyright © 2026 the author, Page 86
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4, No 1, November 2026, page 81-91
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

treated as translation of content rather than as a property of the whole interface, with interface strings left in the development language. This is a common failure in bilingual applications because developers who are fluent in both languages do not experience mixed labels as a barrier. A similar developer-centred view explains the voice-input issue: the transcript was the output the system needed, but the user needed to know that the system was listening, a basic requirement of visibility of system status (Nielsen, 1994). In both cases the problem was invisible from the development perspective and became visible only when the interface was judged on behalf of users whose language and literacy differ from those of the team.

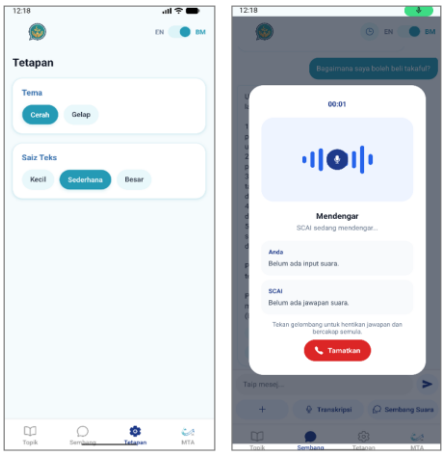
All interface strings were subsequently localised so that switching language changes every label, button and topic text. In BM, the navigation labels now read "Topik", "Sembang", "Tetapan" and "MTA", and the chatbot actions read "Dengar", "Jelaskan Lagi" and "Tamatkan Sesi" (Figures 3b and 4b). Voice input now displays an animated waveform that makes the listening state visible (Figure 3b). A new Settings tab provides light and dark themes and three text sizes (Figure 5a), in line with accessibility guidance that content should remain usable when text is resized and contrast is adjusted (W3C, 2023).

3.4 Conversational Interaction Enhancement

The chatbot was considered familiar and user-friendly because its layout follows the conventions of modern generative AI chat applications, and the conversation layout, prompt area and bilingual flow were judged clear. The experts and client representatives, however, identified gaps in how the chatbot introduced itself and supported its answers. First-time users were given no explanation of what the chatbot could do, responses did not show where the information came from, and the language of responses was not always consistent with the language selected in the application. The panel further suggested embedding external links to additional resources, supporting voice-to-voice conversation and adding explainer video placeholders to the topic overlays to enrich the educational content. These recommendations were mapped to the BUS constructs of the ease of starting a conversation and the visibility of and access to chatbot functions (Borsci et al., 2022).

These gaps arose because the prototype adopted the interaction model of general-purpose generative AI chat, in which an empty prompt box invites any question and the assistant is presumed to need no introduction. That model presupposes users who already know what such systems can do; first-time and lower-literacy users lack this mental model and tend to form inaccurate expectations of what an agent can do (Luger & Sellen, 2016). Similarly, although SCAI drew on a curated knowledge base, the interface presented each answer as a finished product and hid the evidence behind it, and the response language was not bound to the application setting. In a general assistant these omissions are minor; in a *Shariah*-compliant agent, where a fluent but unsupported answer could misinform users about religious obligations (Ji et al., 2023), they directly affect whether guidance can be trusted. The domain experts' insistence on visible sources and links to MTA resources reflects this higher standard of accountability.

The refined chatbot opens with a greeting that explains its purpose and capabilities, presents multi-step answers as numbered lists, displays a source label beneath each response and inserts external links where further reading is available (Figure 3b). Responses are generated in the language currently selected in the application. A voice-to-voice mode was added that allows users to speak to SCAI and hear its spoken reply through a dedicated screen that shows both the user's input and the agent's response (Figure 5b), and video placeholders were embedded in the topic overlays (Figure 2d).



(a) Settings tab (new) (b) Voice-to-voice (new)

Figure 5. New settings tab and voice-to-voice interaction screen

3.5 Traceability of Refinements

Table 4 consolidates the link between each theme, its construct, the experts' recommendations and the implemented refinements. Each recommendation recorded in the review was mapped to at least one refinement in the final prototype, and every refinement can be traced back to a specific theme and usability construct. Presenting the outcome in this form

Copyright © 2026 the author, Page 87
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4, No 1, November 2026, page 81-91
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

makes the review auditable: a reader can see not only what was changed but which expert concern and which established usability dimension motivated the change.

Table 4. Traceability of expert recommendations to implemented refinements

Theme	Construct	Expert recommendation	Implemented refinement	Figure
Visual design and branding	QUIS	Apply MTA colour theme, strengthen branding and visual hierarchy, simplify logo	MTA blue theme, low-opacity Islamic decorative background, emblem-only header logo, tinted topic cards	2c, 2d
Navigation and interface interaction	QUIS	Improve close control and overlay size, separate chat input functions, add settings and MTA access	Enlarged overlay with button-shaped close control, two-row chat input, four-tab navigation with Settings and MTA	2d, 3b, 4b
Language consistency and accessibility	USE	Localise all interface elements, visualise voice transcription, add font and display options	Full BM/EN localisation, animated voice waveform, text size options, light and dark modes	3b, 4b, 5a
Conversational interaction enhancement	BUS	Introduce chatbot capabilities, show sources, keep response language consistent, add links, voice and video	Greeting message, numbered answers, source labels, external links, language-consistent responses, voice-to-voice mode, video placeholders	2d, 3b, 5b

3.6 Complementarity of the Two Expert Lenses

The dual-expert design produced complementary rather than redundant findings, and the difference can be characterised as two lenses. The UI/UX experts applied an operability lens: they asked whether users could see, reach and act on interface elements, and concentrated on the visibility of controls, the layout of the input area, navigation discoverability and feedback during voice input. The content experts applied a legitimacy lens: they asked whether the application said the right things, in the right tone, on behalf of the right institution, and foregrounded institutional identity, educational presentation and appropriateness for the *Takaful* context. Neither lens alone would have produced the full set of refinements in Table 4. The one point of convergence, branding, is revealing: a concern that usability frameworks classify as aesthetics was raised as a matter of trust, which suggests that in *Shariah*-sensitive domains the boundary between interface quality and content credibility is weaker than existing instruments assume. The findings also qualify the reliance on user-based testing reported for mobile learning (Kumar et al., 2024). Expert inspection surfaced problems such as incomplete localisation and missing source attribution before any end user was exposed to them, problems that novice users might experience as confusion, or as their own failure, without being able to articulate the cause.

3.7 Design Principles for Shariah-Sensitive Conversational Interfaces

Moving beyond the specific fixes, the root-cause analysis in Sections 3.1 to 3.4 points to four design principles that can generalise to conversational agents in *Shariah*-sensitive and other trust-sensitive domains (Table 5). Each principle responds to an issue that existing usability constructs either treat as cosmetic or do not capture.

Table 5. Design principles for Shariah-sensitive conversational interfaces derived from the dual-expert review

Principle	Definition	Evidence from the SCAI review	Relation to existing constructs
Legitimacy cues	Visible signals of the institution and faith context that stand behind the content	Monochrome prototype with weak MTA identity; MTA colour theme, emblem and Islamic motifs adopted	QUIS treats colour as aesthetics; here it acts as a credibility cue (Fogg et al., 2003)
Provenance transparency	Each answer shows where its content comes from and links to authoritative resources	No sources shown in initial chatbot; source labels and MTA links added	BUS addresses response quality but not traceability of sources; mitigates hallucination risk (Ji et al., 2023)
Linguistic completeness	Every label, prompt and response follows the language the user has selected	Mixed BM/EN labels and replies; full localisation and language-bound responses	USE ease of use assumes a single language; partial localisation excludes single-language users
Expectation setting	The agent declares its purpose, scope and limits at first contact	No introduction in initial chatbot; greeting explaining purpose and capabilities	Extends BUS ease of starting a conversation to the scope of Shariah guidance (Luger & Sellen, 2016)

Copyright © 2026 the author, Page 88
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Journal of SAINTEKS
Vol 4, No 1, November 2026, page 81-91
ISSN 3030-8011 (Media Online)
Website https://prosiding.seminars.id/sainteks

The principles of expectation setting and provenance transparency correspond closely to what the chatbot literature identifies as determinants of a good experience. Borsci et al. (2022) included the ease of starting a conversation and the visibility of chatbot functions in BUS precisely because conventional instruments overlook them, and the absence of an introductory greeting in the initial SCAI prototype is an instance of that gap. Rapp et al. (2021) and Følstad et al. (2021) emphasise that users form expectations of a chatbot early and that transparency about its capabilities and limitations shapes trust. In SCAI this concern is amplified, because users must be able to believe that the guidance they receive is consistent with *Shariah* principles. Displaying the source of each answer and linking to MTA resources therefore functions not only as a usability improvement but also as a trust mechanism, which is consistent with the emphasis that reviews of educational chatbots place on response quality and credibility as preconditions for sustained use (Hwang & Chang, 2023; Labadze et al., 2023).

The principle of linguistic completeness addresses a risk that is easy to overlook in bilingual applications. A language switch that changes the main content but leaves buttons or tabs in another language gives the impression of a finished feature while still disadvantaging users who depend on one language. Because the target users of SCAI include marginalized and lower-income communities whose financial and digital literacy varies (Munisamy et al., 2022; Busari, 2025), complete localisation, adjustable text size and visible voice feedback are not cosmetic refinements but conditions for inclusion. Reviews of educational chatbots likewise stress that interaction design must accommodate learners with different needs and preferences (Wollny et al., 2021; Kuhail et al., 2023).

Taken together, the principles suggest a refinement of how usability is understood in this domain. Existing instruments, including BUS, treat usability and trust as separate outcomes, with trust following from good performance. In a *Shariah*-compliant agent the relationship is more direct: an unexplained chatbot, an unattributed answer or a half-translated screen is read not only as a usability flaw but as a reason to doubt whether the guidance is *Shariah*-compliant. Usability thus functions as part of the user's evidence for *Shariah* assurance. This implies that evaluation instruments for such agents need constructs for legitimacy and provenance in addition to conventional usability dimensions, and that domain experts are needed in design review not merely to check content but to define what trustworthy interaction means for the target community.

3.8 Triangulation with End-User Evaluation

Table 6 compares the expert themes with the results of the summative evaluation of the refined application with 33 users (Annamalai et al., 2026). Interface and Interaction Quality, the construct most closely related to the visual design and navigation themes, obtained the highest mean of the six constructs (M = 3.66 on a 0 to 4 scale), and Accessibility and Trust, related to the language and conversational themes, obtained a mean of 3.60, with participants highlighting bilingual support, readable text and trustworthy, respectful responses. The end-user evidence therefore converges with the expert findings: the areas the experts targeted were among those users rated most strongly, and the features users singled out in their comments correspond to refinements the experts recommended. Because the initial and refined prototypes were not compared with users, these results cannot be attributed causally to the expert review. Industry adoption of chatbots in *Takaful* depends on how consumers perceive them (Mohd Thas Thaker et al., 2026), and the convergence suggests that branding, transparency and language consistency are among the design levers that developers can act on before deployment.

Table 6. Triangulation of expert themes with end-user evaluation of the refined application

Expert theme	Related end-user construct	End-user evidence after refinement (n = 33)
Visual design and branding	Interface and Interaction Quality	M = 3.66 (highest of six constructs; 0 to 4 scale)
Navigation and interface interaction	Interface and Interaction Quality	M = 3.66 (highest of six constructs; 0 to 4 scale)
Language consistency and accessibility	Accessibility and Trust	M = 3.60; bilingual support and readable text highlighted in comments
Conversational interaction enhancement	Accessibility and Trust	M = 3.60; trustworthy and respectful responses highlighted in comments

Note: End-user data are from Annamalai et al. (2026) and refer to the refined application only; no comparison with the initial prototype was made.

3.9 Limitations

Several limitations should be noted. Although the panel size is justified for formative inspection (Section 2.6), a small panel cannot claim that all problems were found, given the evaluator effect (Hertzum & Jacobsen, 2001). The review did not include independent *Shariah* advisers or representatives of the target community as separate expert groups, and the thematic analysis was carried out by the project team, which may introduce interpretive bias. The content expert group evaluated the prototype after a guided demonstration rather than through independent use, so their feedback may reflect the developer's framing of the application. Finally, end-user triangulation relied on post-refinement ratings only, so it corroborates but does not test the effect of the refinements.

Copyright © 2026 the author, Page 89
This Journal is licensed under a Creative Commons Attribution 4.0 International License

4. CONCLUSION

This study set out to determine how domain and usability expertise can be combined to shape a *Shariah*-compliant conversational agent for marginalized communities before it reaches them. Its central finding is that the two forms of expertise are not interchangeable: UI/UX experts judged whether the interface could be operated, while *Takaful* domain experts judged whether it could be trusted, and several problems, such as incomplete localisation, unattributed answers and an agent that did not explain itself, mattered precisely because they sat at the intersection of the two. In a *Shariah*-sensitive domain, usability is therefore part of how users judge *Shariah* assurance, and the four principles proposed here express that relationship in a form designers can act on. The findings carry practical implications for the *Takaful* industry. For operators developing customer-facing or educational AI, a dual-expert review involving both *Shariah* or domain experts and UX specialists should be treated as a release gate rather than an optional usability check, and a construct-mapped traceability matrix offers a low-cost way to document that each concern was addressed. Provenance should be designed in from the start: every chatbot answer should be drawn from a *Shariah*-vetted knowledge base and display its source, so that accountability for religious guidance remains visible. Complete localisation, verified through an audit of every interface string, should be an acceptance criterion for applications serving multilingual or marginalized communities. At the policy level, industry bodies such as MTA are well placed to translate these principles into sector guidelines for AI conversational agents, aligning interface requirements with existing *Shariah* governance and consumer protection expectations. Future research should pursue four directions. First, the proposed principles should be operationalised as a *Shariah*-sensitive extension of the Chatbot Usability Scale, with items developed through a Delphi study with *Shariah* advisers, *Takaful* practitioners and HCI experts and validated through confirmatory factor analysis with end users. Second, a controlled experiment with marginalized users comparing the initial and refined prototypes is needed to estimate the causal effect of expert-driven refinements on task success, trust calibration and *Takaful* knowledge gain. Third, the *Shariah* accuracy of generated responses should be audited systematically by *Shariah* advisers and related to user trust, to establish whether provenance cues help users distinguish reliable from unreliable answers rather than simply increasing trust. Fourth, replication in other Muslim-majority markets, such as Indonesia and Brunei, would test whether the principles hold across different regulatory, linguistic and cultural contexts.

5. ACKNOWLEDGEMENT

The authors thank the Malaysian Takaful Association (MTA) and the expert reviewers for their constructive feedback on the SCAI prototype. This research received Social Innovation (Industry) funding from FWD Insurance and Family Takaful Company (S/O Code: 22185).

REFERENCES

Annamalai, S., Kuo, W. S., Che Omar, A., Yusof, N., & Abd Muin, M. A. (2026). *SCAI: User acceptance of a Shariah-compliant conversational agent for promoting Islamic financial literacy among marginalized communities* [Manuscript submitted for publication]. School of Multimedia Technology and Communication, Universiti Utara Malaysia.

Borsci, S., Malizia, A., Schmettow, M., van der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2022). The Chatbot Usability Scale: The design and pilot of a usability scale for interaction with AI-based conversational agents. *Personal and Ubiquitous Computing, 26*(1), 95–119. https://doi.org/10.1007/s00779-021-01582-9

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Busari, S. A. (2025). Maqasid al-Shariah-based strategies for enhancing digital adoption among B40 Muslim entrepreneurs. *Journal of Islamic Finance, 14*(1).

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum, 41*(5), 545–547. https://doi.org/10.1188/14.ONF.545-547

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 213–218). ACM. https://doi.org/10.1145/57167.57203

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education, 40*(4), 314–321. https://doi.org/10.1111/j.1365-2929.2006.02418.x

Elwy, A. R., Wasan, A. D., Gillman, A. G., Johnston, K. L., Dodds, N., McFarland, C., & Greco, C. M. (2020). Using formative evaluation methods to improve clinical implementation efforts: Description and an example. *Psychiatry Research, 283*, 112532. https://doi.org/10.1016/j.psychres.2019.112532

Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003). How do users evaluate the credibility of Web sites? A study with over 2,500 participants. In *Proceedings of the 2003 Conference on Designing for User Experiences* (pp. 1–15). ACM. https://doi.org/10.1145/997078.997097

Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., & Luger, E. (2021). Future directions for chatbot research: An interdisciplinary research agenda. *Computing, 103*(12), 2915–2942. https://doi.org/10.1007/s00607-021-01016-7

Hertzum, M., & Jacobsen, N. E. (2001). The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction, 13*(4), 421–443. https://doi.org/10.1207/S15327590IJHC1304_05

Hwang, G.-J., & Chang, C.-Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments, 31*(7), 4099–4112. https://doi.org/10.1080/10494820.2021.1952615

This Journal is licensed under a Creative Commons Attribution 4.0 International License

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), Article 248. https://doi.org/10.1145/3571730

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). SAGE Publications.

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies, 28*(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Kumar, B. A., Chand, S. S., & Goundar, M. S. (2024). Usability testing of mobile learning applications: A systematic mapping study. *International Journal of Information and Learning Technology, 41*(2), 113–129. https://doi.org/10.1108/IJILT-03-2023-0029

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education, 20*, Article 56. https://doi.org/10.1186/s41239-023-00426-1

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Luger, E., & Sellen, A. (2016). "Like having a really bad PA": The gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5286–5297). ACM. https://doi.org/10.1145/2858036.2858288

Lund, A. M. (2001). Measuring usability with the USE questionnaire. *Usability Interface, 8*(2), 3–6.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382–385. https://doi.org/10.1097/00006199-198611000-00017

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research, 26*(13), 1753–1760. https://doi.org/10.1177/1049732315617444

Medhi, I., Patnaik, S., Brunskill, E., Gautama, S. N. N., Thies, W., & Toyama, K. (2011). Designing mobile interfaces for novice and low-literacy users. *ACM Transactions on Computer-Human Interaction, 18*(1), Article 2. https://doi.org/10.1145/1959022.1959024

Mohd Thas Thaker, H., Lacheheb, Z., Lechumanan, B., Duasa, J., & Taghizadeh-Hesary, F. (2026). Examining factors influencing chatbot adoption in Takaful industry: Malaysian evidence. *International Journal of Islamic and Middle Eastern Finance and Management, 19*(3), 614–637.

Munisamy, A., Sahid, S., & Hussin, M. (2022). A financial literacy model of Malaysian B40 households: The case of financial well-being, education level and socioeconomic status. *Proceedings, 82*(1), 64. https://doi.org/10.3390/proceedings2022082064

Nawang, W. R. W., Johari, F., Shukor, S. A., Shahwan, S., & Abdullah, M. (2025). Islamic financial literacy in Malaysia: Evidence from a young Muslim population. *Global Business and Management Research, 17*(3), 127–138.

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 25–62). John Wiley & Sons.

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (pp. 206–213). ACM. https://doi.org/10.1145/169059.169166

Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An initial model of trust in chatbots for customer service: Findings from a questionnaire study. *Interacting with Computers, 31*(3), 317–335. https://doi.org/10.1093/iwc/iwz022

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence, 2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033

Osman, I., Syed Alwi, S. F., Rehman, M. A., Muda, R., Hassan, F., Hassan, R., & Abdullah, H. (2024). The dilemma of millennial Muslims towards financial management: An Islamic financial literacy perspective. *Journal of Islamic Marketing, 15*(1), 59–78.

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. https://doi.org/10.1002/nur.20147

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies, 151*, 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

W3C. (2023). *Web Content Accessibility Guidelines (WCAG) 2.2*. World Wide Web Consortium. https://www.w3.org/TR/WCAG22/

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence, 4*, 654924. https://doi.org/10.3389/frai.2021.654924

Copyright © 2026 the author, Page 91
This Journal is licensed under a Creative Commons Attribution 4.0 International License