Human Factors in Accessibility, Assistive Technology and Digital Environments, Vol. 202, 2026, 40–49
https://doi.org/10.54941/ahfe1007295

Blind and Low Vision Users' Experience With AI-Infused Banking Chatbots: AI-Specific Experience Dimensions and System Usability

Ekin Sözer and Zeynep G. Karapars

Özyeğin University, 34794 Istanbul, Türkiye

ABSTRACT

Artificial intelligence (AI) infused chatbots have the potential to radically transform the user experience with digital banking for blind and low vision (BLV) users. AI chatbots offer a radically different interaction style. A considerable part of the theories and principles developed for conventional interaction styles becomes less directly applicable. There is limited empirical research on BLV users' experience (UX) with AI-infused banking chatbots. This study aims to explore the AI-specific UX dimensions as well as the conventional UX dimensions of banking chatbots regarding BLV users. The empirical study includes a usability test, a semi-structured interview, and a perceived usability measurement scale. The participants are selected from BLV adults who actively use digital banking services. Findings reveal that perceived intelligence is closely intertwined with accessibility performance. Failures in screen reader compatibility, navigation clarity, state feedback, and table interpretation not only reduced task success but also diminished trust in the system. Text-heavy chatbot structures supported simple, well-defined tasks more effectively, whereas graphical-heavy designs caused navigational breakdowns despite appearing accessible. Moreover, financial interactions were strongly associated with independence, control, and dignity, highlighting the existential significance of accessibility beyond technical compliance. This study aims to contribute to the literature by identifying AI-specific user experience dimensions in the BLV context, and also to provide designers in the digital banking sector with applicable design recommendations for developing better chatbot interactions for BLV users. This paper presents the preliminary findings of a study that is still ongoing.

Keywords: Artificial intelligence, Chatbots, User experience, Accessibility, Blind and low vision, Digital banking, Human-centered AI

INTRODUCTION

Access to digital media and technologies is one of the main research topics in Blind and Low Vision (BLV) accessibility studies (Medina et al., 2025). Within this domain, financial autonomy remains a critical challenge. Being able to manage one's money is an essential part of day-to-day tasks, yet individuals with sensory or functional impairments often face structural constraints that limit their independence in financial decision-making (Tilse et al., 2007).

Received February 20, 2026; Revised March 31, 2026; Accepted April 17, 2026; Available online July 20, 2026

Specifically, BLV users continue to encounter significant barriers in digital banking. Prior research indicates that online banking platforms and self-service technologies are frequently not designed with accessibility in mind, which limits independent use through assistive technologies (Hassan, Abd El-Aziz and Hamza, 2020). These barriers are not only technical but also institutional, as complex procedures and insufficient support mechanisms can lead to exclusion and discrimination (Nadela and Yulianti, 2022). Furthermore, interaction design flaws in mobile applications, such as dense visual elements, poor labeling, and complex task flows, hinder screen reader performance and directly impact the privacy, security, and emotional well-being of BLV users (Nadela and Yulianti, 2022; Kerdar, Bächler and Kirchhoff, 2024). Based on a literature survey on research about Blind and Low Vision, Medina et al. (2025) suggest that recent developments in AI represent a potential research trajectory, as AI is highly promising for BLV people. In the last few years, AI technology has been increasingly adopted by traditionally tech-cautious banking services, first in call centers as customer assistants and more recently in mobile applications as chatbots. These AI chatbots have the potential to bypass the aforementioned accessibility barriers by providing more intuitive interfaces, thereby increasing BLV users' safety from fraud and fostering independence in managing personal finances. However, there are relatively few studies on BLV user experience with AI chatbots, and even fewer that focus specifically on personal banking applications.

Hence, we conducted an empirical study to explore the BLV user's experience with AI chatbots in mobile banking applications. AI-infused products, such as chatbots, involve interactions that differ considerably from conventional graphical user interfaces (GUIs). Therefore, the dimensions of user experience (UX) regarding AI-infused products are emerging as new research is undertaken. Spallazzo et al. (2025) define UX dimensions with AI-infused products as trustworthiness, which includes reliable performance, data management, human oversight, privacy, and transparency. Secondly, they define intelligence, characterized by accuracy, adaptability, and context-awareness. The third dimension is conversationality, which refers to voice quality and natural language. Trustworthiness, alongside safety and reliability, is also emphasized by Shneiderman (2020). Regarding voice user interfaces (VUIs), Klein et al. (2023) suggest dimensions including response behavior, response quality, and comprehensibility. Spallazzo et al. (2025) also found that conventional UX dimensions, including meaningfulness, hedonic aspects, and pragmatic aspects, remain applicable to AI-infused products. Meaningfulness is a cognitive process where users interpret interactions and assign symbolic significance to products, leading to emotional responses (Desmet and Hekkert, 2007). The hedonic dimension, originally from the marketing discipline (Voss et al., 2003), relates to joy, fun, thrilling aspects of experience. The pragmatic dimension focuses on reaching goals efficiently and effectively (Bevan, 2008). In this study, we based our analysis on the dimensions defined above to explore how the BLV users experience AI-infused mobile banking chatbots with different design strategies.

METHODOLOGY

This study adopts a qualitative-dominant mixed-methods design to explore the interactions of blind and low vision (BLV) individuals with AI-infused banking chatbots, focusing on the dimensions of user experience before, during and after these interactions. The research is conducted as a field study in which participants engage with real-world mobile banking interfaces.

Participants

The participant group consists of five BLV individuals working at Istanbul Dialogue Museum who volunteered to participate. Inclusion criteria were being legally blind or low vision, being age 18 or older, actively using smartphone with assistive technologies, and being a regular user of at least one mobile banking app. Four out of five participants had no prior experience with banking chatbots. Ethical procedures were followed with informed consent read aloud and verbal consent recorded.

Setting and Apparatus

Data collection sessions took place at Istanbul Dialogue Museum, participants' daily workplace. To protect financial and personal data privacy, participant's personal accounts and mobile phones were not used. They used researcher-owned bank accounts with limited real balances. Hardware/software were Apple iPhone 11 with VoiceOver and Xiaomi Redmi Note 14 with TalkBack. One Android user opted for the external screen reader Jieshuo+, configured with researcher support.

Procedure

Each data collection session lasted approximately 60 minutes and consisted of four structured phases. First, a brief preliminary interview was conducted to explore demographics, visual status, educational background, and access to banking services. Secondly, formative usability testing was performed where participants completed four tasks using chatbots in mobile applications of three banks. The banks were chosen according to the design strategy in their AI-infused chatbots. One of the banks was a leading public bank and the other two were major private banks in Türkiye.

The Scenarios in the usability test were: 1. Learning the current gold exchange rate, 2. Checking the remaining credit card limit, 3. Reviewing daily expenditures, and 4. Sending a rent payment to a registered recipient with a specific transaction description, "rent". To minimize learning and fatigue effects, the order of bank testing was counterbalanced among participants, ensuring each followed a unique sequence for an unbiased comparison. Thirdly, after testing each banking chatbot, the Turkish adaptation of 10-item System Usability Scale (SUS) (Demirkol and Şeneler, 2018; Brooke, 1996) was administered to assess perceived usability. The researcher read the questions and scale options aloud, and participants provided ratings on a 5-point Likert scale. Finally, a semi-structured interview was conducted to gather in-depth insights on the user's chatbot experience.

Data Analysis

All collected data including audio recordings, screen videos and SUS survey responses were subjected to a multi layered analysis process. Audio recordings were transcribed verbatim. Video recordings were analyzed and hand-written into the transcriptions as observation notes. For qualitative analysis two independent researchers opened separate project files in Atlas.ti (v9.24.3) qualitative analysis software and performed initial coding on the same dataset. Thematic analysis followed the inductive approach proposed by Braun and Clarke (2006). One of the researchers has a psychology background, the other has an interaction design and usability background. Due to the differences in expertise, the first researcher focused on emotional and immersive user experience and the second researcher focused on design and usability aspects. Therefore, a consensus based coding process was adopted. Rather than measuring inter coder reliability the aim was to surface diverse interpretative perspectives. After doing initial coding individually, the coders met several times and compared their outputs, merged conceptually similar codes, clarified definitions and jointly developed a shared codebook. Subsequently these unified codes were deductively mapped onto the AI specific user experience dimensions suggested in the literature. Placement of each code into these dimensions was determined through mutual agreement. Quantitative data from SUS scores were analyzed using descriptive statistics to support qualitative findings and quantify differences in usability perceptions across chatbots.

RESULTS

The results are presented under three main headings: participant profile, task success rates, System Usability Scale (SUS) scores, and qualitative thematic analysis results.

Participant Profile

The participant group consists of 5 individuals (4 Blind, 1 Low Vision) aged 25, 31, 38, 45, and 53, with varying levels of education.

Table 1: Participant profile (N = 5).

Variable	Category	n
Vision Status	Blind	4
	Low Vision	1
Education	Secondary	1
	Associate	1
	University	3
Device	iPhone / VoiceOver	4
	Android / Jieshuo	1

Usability Testing Findings

Task success rates across four different scenarios (S1: Gold exchange rate S2: Credit card limit S3: Credit card spending S4: Rent payment) was categorized as shown in Table 2.

Table 2: Task success rates (%) by banking chatbot and scenario.

Scenario	Bank A	Bank B	Bank C
S1. Gold exchange rate	60%	100%	80%
S2. Credit card limit	80%	100%	0%
S3. Credit card spending	40%	40%	40%
S4. Rent payment	0%	60%	0%

During the interview, three users said they found the Bank C chatbot very easy to use. However, usability test results showed that it had the highest failure rate. Therefore, the actual results and user perceptions were not consistent. The interface looked like it was easy to use, but in reality it was not effective. This phenomenon can be explained by the high accessibility level of Bank C, because there were very few missing tags or alt-text, and the tables were designed with accessibility in mind. Users did not encounter the usual accessibility issues, since it is a hygiene factor, they were satisfied with the interface even though it was not the most effective one.

Discussion on SUS Application in BLV Context

SUS scores applied to measure perceived usability levels are summarized in Table 3.

Table 3: System usability scores.

Banking App	n	mean	sd	median	min	max
SUS- A	5	50	29,8	47,5	22,5	97,5
SUS- C	5	61,5	34,6	65	25	100
SUS- B	5	70,5	33,1	77,5	17,5	100

Qualitative observations revealed that the SUS scale has significant methodological limitations for BLV users. Participants frequently preferred extreme ratings and struggled with reverse-scored items, likely due to the cognitive load of evaluating questions read aloud. Frequent inconsistencies between verbal feedback and numerical scores further undermined data validity. Consequently, SUS results were not treated as a reliable standalone indicator but were interpreted alongside qualitative findings.

Findings of the Thematic Analysis

This section presents the findings from usability tests and interviews conducted with BLV participants about AI-infused banking chatbots. The findings are categorized under the themes of Pragmatic Aspects, Intelligence, Conversationality and Meaningfulness.

Pragmatic Aspects: The banking chatbots in this study employed design strategies ranging from "text- heavy" to "graphical element heavy" design. Graphical-heavy chatbots, utilized textual, human language sentences mostly at the beginning of the response and then presented visually designed buttons and menu links in the response area. Those buttons and menu links directed the user outside the chatbot, into the relevant 'location' in the application. Text heavy chatbots generated textual response with minimal menu links and buttons, and the user was rarely directed outside the chatbot. Dialogue heavy interaction was found more practical by two users as it eliminated the necessity of reading all headings, menus and buttons with a screen reader. Also, the success rate of all users in the usability test was the highest with the text heavy chatbot (Bank B). On the other hand, it was also observed that textual heavy chatbots could produce irrelevant responses when the user's goal was not well-defined or the task was a complex, multi-step one. Conversely, GUI heavy interaction was perceived more useful for complex and ill-defined goals, i.e., Scenario 4: paying the rent, since the AI chatbot provided relevant options by listing alternative menus in the response area. Consequently, it was determined that, given the current intelligence level of the AI, text based interaction design strategy is more usable for simple and specific tasks, while GUI based structures are more usable for complex and ill-defined tasks. One strategy mentioned by all users were memorising the layout of screens and information architecture of the interfaces (of their banking applications). It is observed that they tried memorisation strategy at the beginning of the tests. However, the dynamic and ephemeral nature of chatbots rendered this strategy very hard, if not impossible. Efforts to memorize dynamic chatbot interfaces especially those heavy with graphical elements resulted in more errors, and sometimes failure of the tasks. The most frequently used chatbot interface functions, such as composing a prompt, sending the prompt, and reading the response, were visually prioritised by the help of graphical elements and default settings of the cursor. On the other hand, BLV users had to manually search for the input field, the send button and the beginning location of the latest response every time. Audio or gestural short-cuts for the above mentioned functions can significantly enhance the efficiency of the chatbot interface for BLV users. The state of all the chatbots were unclear for BLV users; there were no auditory or tactile indicators for chatbot states such as "ready for prompt input", "generating response" and "response is ready". Similarly there was a lack of auditory or tactile feedback when a prompt was sent. Users waited for a few seconds after entering a prompt, then started searching for the response. A significant navigational error occurred when the cursor was incorrectly positioned at the end (instead of at the beginning) of the chatbot's response, misleading users to return to older responses at the top of the screen. This caused users to conclude that the chatbot had either provided an incorrect response or failed to respond entirely. Last but not least, being able to 'see' all the control and display elements, headings, content text, images and animations in a user interface is the hygiene factor for accessibility of BLV users. During the tests, it was observed that screen readers such as VoiceOver and TalkBack failed to identify tags for buttons, links, and icons within the chatbot interface. One of the major consequences of this type of accessibility problem is that users could not recognise the borders of the chatbot interface, so they got out of the chatbot and had difficulty getting back in. Another obstacle was the marketing oriented expressions in the headings within the interface, which decreased the users' ability to distinguish interface components from content. These two would not be an issue for non-BLV users, because the chatbot borders and headings were signified by visual design, so it was easy to distinguish them. Table accessibility is crucial in banking applications, as financial information is usually presented with tables, for example credit card expenditure, personal account balance, investment fund rates, etc. It is observed that the screen readers scanned the tables horizontally, whereas the information in the tables were organized vertically, as also pointed out by Dix et al. (p. 374, 2004). This caused users to get incorrect information.

Intelligence: In examining intent recognition, context awareness, and complex task management, it was observed that the systems do not yet possess the capacity to plan and execute multi-step tasks. When faced with a complex objective, the chatbot typically directed the user to the first step of the task but failed to provide guidance for subsequent steps. For instance, in Scenario 4, which required rent payment to a registered account with an added description, users struggled to complete the process because the chatbot's guidance ceased after they reached the registered account. One user explicitly stated uncertainty regarding how to proceed at this stage. Users expressed that their expectation from an AI is more conversation based guidance. Performance varied significantly regarding context awareness and computational capacity. In Scenario 2, which required calculating the remaining credit card limit by subtracting expenses, the GUI heavy structure of Bank C resulted in failure for all users. While Bank A could not perform the calculation, it provided a "relevant but incomplete" response by listing the total expenses. Bank B was the only AI chatbot capable of performing the calculation and providing a correct, one sentence answer. These failures altered the interaction strategies of the users. For example, one user began entering prompts with a prior assumption of failure and low motivation in subsequent tasks. Another user initiated "prompt restructuring" by breaking tasks into smaller sub-components to ensure the AI chatbot could understand. Regarding questions about current gold prices or daily credit card expense lists (Scenarios 1 and 3), chatbots were observed to present raw data tables or unfiltered lists instead of providing direct, specific answers. Although Bank C provided an accessible table, users expressed dissatisfaction, noting that the AI failed to understand their intent and did not interpret the data to produce a specific response, such as the selling rate of gold.

Conversationality: It was determined that chatbots heavily utilize banking terminology and technical language. Even when users entered the prompts in daily language and the AI retrieved correct information, the lack of human language for these information pieces prevented users from identifying if they accomplished their goals in the scenarios. The case about technical language was encountered when users asked if there were any expenditures made with the credit card. The AI chatbot's response was "no operation found for your criteria". The response was technically correct because no expenditures were made actually. However, users thought the AI could not find the course of action for that query. Such experiences led users to enter shorter, keyword-like prompts and utilize technical language whenever they can. Analysis of interaction strategies of users revealed that users initially formed sentences in natural language but as there were breaches in conversational quality of AI, they resorted to technical-sounding language.

Meaningfulness: All participants emphasized that mobile banking applications are critical for their quality of life in terms of financial security, privacy, and autonomy. However, the corruption of accessibility settings (tags, alt-texts, etc.) with every update and the subsequent user-need to restart the learning process is a source of severe frustration. At this juncture, AI chatbots are viewed as tools with the potential to bypass the accessibility issues inherent in graphical user interfaces.

DISCUSSION

The technical issues observed in our study, such as screen readers skipping buttons or misreading tables, directly harmed the user experience. These failures made tasks harder and led users to perceive the chatbot as lacking intelligence. When a system fails to convey information clearly, users begin to question its overall competence, which ultimately results in a loss of trust. As Metatla et al. (2020) emphasize, accessibility should not be an afterthought or a secondary feature. Instead, technology must be inclusive from the outset because assistive tools like screen readers only function effectively when the underlying interface structure is accessible. Participants often interpreted interaction failures as a personal inability to choose the right words. This behavior corroborates the concept of design-induced self-blame proposed by Norman (2013). From the perspective of situated action (Barnick, 2009), these issues arise because systems adhere rigidly to predefined scripts instead of adapting to immediate social contexts. Current banking chatbots often fail to grasp situational intent or linguistic variations, which causes frequent interaction breakdowns. Consequently, users internalize these technological limitations as personal failures. Their feelings of guilt stem directly from the AI's lack of contextual intelligence. Therefore, the solution lies in enhancing the system's mutual intelligibility to respond more effectively to human context rather than training the users themselves. One participant highlighted the insulting nature of being dependent on others, viewing the chatbot as a potential way to break this cycle of need. However, this hope for independence is frequently hindered by interfaces designed primarily for sighted users. Reproachful expressions from participants, such as "this is not for me," indicate that poor accessibility is more than a technical flaw. It is a factor that creates a sense of exclusion and secondary citizenship. As Spallazzo et al. (2025) note, meaningfulness in products arises when they convey a clear purpose and personal significance. When accessibility is lacking, this meaningful connection is disrupted. This prevents the product from fulfilling its potential to create an empowering experience for all users.

CONCLUSION

This study shows that AI-infused banking chatbots can greatly support the financial independence of blind and low vision users. However, current designs often fail to meet the specific needs of this user group. As noted by Power et al. (2012) accessibility guidelines are not enough by themselves. Real user testing is necessary to find and fix actual usability problems. Our observations confirm common issues such as unclear feedback poor guidance and confusing information structure. These problems mainly affect BLV users because visual design principles that work for sighted users do not translate well to screen reader experiences. Chatbot interfaces are dynamic. Text changes length and buttons appear within responses. This makes it harder for BLV users to keep track of where they are. Designers should go beyond traditional visuals and adopt multi-modal approaches. Adding non-speech sounds and audio cues can improve usability as suggested by Dix et al. (2004). For example using distinct sounds to show when a response is loading or assigning gestures for common actions like sending a message can reduce effort. Screen readers already use some of these methods. Banking apps should follow these existing patterns to make accessibility a core feature. One helpful design was the audio repetition of user input. This let participants check their commands against background noise. It also helped them feel more confident about what they had typed. Clear separation between the chatbot and the rest of the app is also important. Embedding clickable buttons in chat responses may help sighted users but often confuses BLV users. They may accidentally leave the chatbot and struggle to return. Bank B in our study succeeded because it kept most tasks inside a text-based flow. For AI features we suggest using everyday language instead of banking jargon. AI systems can learn from how people talk. This helps connect user goals with complex financial terms. The chatbots we tested did not make things up. But they did not clearly say when they could not handle a task. Better transparency means telling users what the system can and cannot do. This kind of honesty helps guide people through hard steps. In summary banking chatbots can improve life for BLV users. But only if designers understand how different these conversations are from traditional apps. Success comes from building systems that respect both the limits and strengths of spoken interaction.

ACKNOWLEDGMENT

The authors would like to thank the Istanbul Dialogue Museum and its staff for their voluntary participation and support during the data collection process. This work was also supported by Özyeğin University.

REFERENCES

Barnick, H. (2009) 'Review of Human-Machine Reconfigurations: Plans and Situated Actions by Lucy A. Suchman', Subjectivity, 28, pp. 345–349. doi: 10.1057/sub.2009.11.

Bevan, N. (2008) 'Classifying and selecting UX and usability measures', in COST 294-MAUSE Workshop: Meaningful Measures: Valid Useful User Experience Measurement. June 2008.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', Qualitative Research in Psychology, 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.

Brooke, J. (1996) 'SUS - A Quick And Dirty Usability Scale', Usability Evaluation In Industry, 189(194), pp. 4–7. doi: 10.1002/hbm.20701.

Demirkol, D. ve Şeneler, Ç. (2018) 'Translation of the System Usability Scale into Turkish: SUS-TR', Uşak University Journal of Social Sciences

Desmet, P.M.A. and Hekkert, P. (2007) 'Framework of product experience', International Journal of Design, 1(1), pp. 57–66.

Dix, A., Finlay, J., Abowd, G. and Beale, R. (2004) Human-Computer Interaction. 3rd edn. England: Pearson Education Limited.

Hassan, A., Abd El-Aziz, R. and Hamza, M.A. (2020) 'Barriers to digital banking for the visually impaired', Journal of Financial Services Marketing, 25(1), pp. 1–12.

Kerdar, S.H., Bächler, L. and Kirchhoff, B.M. (2024) 'The accessibility of digital technologies for people with visual impairment and blindness: a scoping review', Discover Computing, 27(24).

Klein, A.M., Kölln, K., Deutschlander, J. and Rauschenberger, M. (2023) 'Design and Evaluation of Voice User Interfaces: What Should One Consider?', in Proceedings of the HCII 2023. Springer, pp. 167–190. doi: 10.1007/978-3-031-35921-7_12.

Medina, N.M., de Castro, R., de Oliveira, R. and de Almeida, N. (2025) 'Blind and Low Vision Users' Interaction with AI: A Systematic Literature Review', in Proceedings of the AHFE 2025.

Nadela, S.D. and Yulianti, L.P. (2022) 'Inclusive design of digital banking with voice user interface: A study based on Indonesia's population', in 2022 International Conference on Information Technology Systems and Innovation (ICITSI). IEEE.

Norman, D.A. (2013) Gündelik Nesnelerin Tasarımı [The Design of Everyday Things]. trans. İ. Kalınyazgan. Ankara: TÜBİTAK Popüler Bilim Kitapları.

Power, C., Freire, A., Petrie, H. and Swallow, D. (2012) 'Guidelines are only half of the story: Accessibility problems encountered by blind users on the web', in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). New York: ACM, pp. 433–442. doi: 10.1145/2207676.2207736.

Shneiderman, B. (2020) 'Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy', International Journal of Human-Computer Interaction, 36(6), pp. 495–504. doi: 10.1080/10447318.2020.1741118.

Spallazzo, D., Sciannamè, M. and Ceconello, M. (2025) User experience + artificial intelligence: Assessing the qualities of ai-infused systems. Springer.

Tilse, C., Setterlund, D., Wilson, J. and Rosenman, L. (2007) 'Managing the Financial Assets of Older People: Balancing Independence and Protection', The British Journal of Social Work, 37(3), pp. 565–572.

Voss, K.E., Spangenberg, E.R. and Grohmann, B. (2003) 'Measuring the hedonic and utilitarian dimensions of consumer attitude', Journal of Marketing Research, 40(3), pp. 310–320.