"I know my data doesn't leave my phone, but still feel like being wiretapped": Understanding (Mis)Perceptions of On-Device AI Vishing Detection Apps

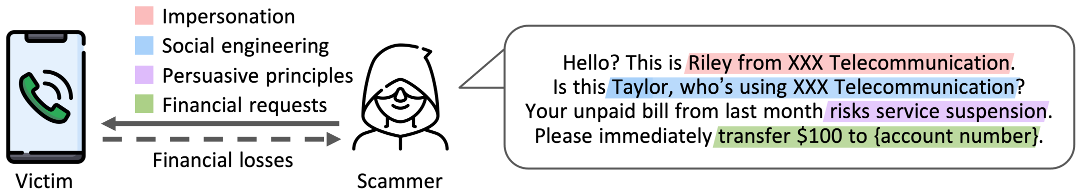
Subin Park
School of Electrical Engineering
KAIST
Daejeon, Republic of Korea
subin.park@kaist.ac.kr

Hyungjun Yoon
School of Electrical Engineering
KAIST
Daejeon, Republic of Korea
hyungjun.yoon@kaist.ac.kr

Janu Kim
School of Computing
KAIST
Daejeon, Republic of Korea
janukim@kaist.ac.kr

Hyoungshick Kim
Sungkyunkwan University
Seoul, Republic of Korea
hyoung@skku.edu

Sung-Ju Lee
School of Electrical Engineering
KAIST
Daejeon, Republic of Korea
profsj@kaist.ac.kr



Hello? This is Riley from XXX Telecommunication.
Is this Taylor, who's using XXX Telecommunication?
Your unpaid bill from last month risks service suspension.
Please immediately transfer $100 to {account number}.

Figure 1: Example of a vishing tactic. A scammer *impersonates* a telecommunication service employee, reveals private information (*social engineering*), and intimidates the victim by using authority (*persuasive principles*) to request *financial transfer*.

Abstract

Vishing, or voice phishing, is a growing global threat exploiting calls to steal sensitive information or money. While on-device AI apps offer promising solutions for real-time vishing detection by analyzing the content of phone conversations, little is known about user perspectives on these tools. To address this gap, we conducted a study with 30 participants using a prototype app featuring on-device AI for speech recognition and vishing detection. We found negligible impacts of on-device AI vishing detection models on smartphone usage satisfaction, but user interviews revealed persistent privacy concerns. Despite the system's use of on-device AI to ensure data security, some participants reported feeling *"being wiretapped."* These findings highlight the need to design privacy-preserving on-device AI solutions and improve user understanding to encourage widespread adoption.

CCS Concepts

• **Human-centered computing** → Usability testing; Laboratory experiments; *Ubiquitous and mobile computing systems and tools*; • **Security and privacy** → **Usability in security and privacy**; **Social aspects of security and privacy**.

Keywords

Usable Security, Vishing, On-Device AI, User Perceptions

ACM Reference Format:
Subin Park, Hyungjun Yoon, Janu Kim, Hyoungshick Kim, and Sung-Ju Lee. 2025. "I know my data doesn't leave my phone, but still feel like being wiretapped": Understanding (Mis)Perceptions of On-Device AI Vishing Detection Apps. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3706599.3719784

1 Introduction

Vishing, or voice phishing, is a rapidly increasing call-based scam leading to significant financial losses worldwide [18, 51, 62]. In 2023, 56.2 million Americans fell victim to vishing, leading to a 25.4 billion USD loss [77]. East Asia faces an even greater challenge, where vishing has become a major cybercrime [20, 24, 37]; in South Korea alone, 19K victims reported 324 million USD in losses in 2023 [1]. Vishing tactics include impersonation [17, 48, 56, 58, 80], deepfake [9, 23, 68, 80], persuasion principles [8, 34, 40], and social engineering [42, 48, 58, 60, 79, 80], often exploiting social crises like the COVID-19 pandemic [15]. Vishing, as described in Fig. 1, utilizes dynamically tailored, realistic conversations, making it challenging for victims to rely solely on intuition to detect them.

CHI EA '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3719784

To combat vishing, blacklist-based systems for blocking suspicious numbers [54, 78, 83, 86], monitoring malicious apps [7], and detecting abnormal permissions, call redirection, or screen overlays [30, 42] have been proposed. Advanced language models have been used to distinguish human and automated callers [52]. However, traditional methods often fail to detect subtle vishing schemes with deceptive conversations.

Recently, artificial intelligence (AI)-based detection has emerged, leveraging contextual understanding of phone conversations [43, 73]. AI vishing detection is effective against complex tactics but raises privacy concerns, especially when sensitive data such as phone conversations is shared with remote servers [33, 84]. However, on-device AI, which processes data locally on the user's device, mitigates these privacy risks. Previous research has integrated on-device AI into mobile security, such as authentication [35, 85] and malware detection [67], primarily focusing on high performance.

While prior studies have explored user evaluations of on-device AI security solutions with a focus on factors such as accuracy, latency [85], and attitudes toward sharing sensor data [35], there is limited understanding of how users comprehensively perceive on-device AI security solutions that require privacy-sensitive data (e.g., phone conversations). Moreover, though security apps are commonly executed in the background, on-device AI demands substantial computational resources, causing performance degradation in other apps and device overheating issues [41]. Given that poor user experience leads to uninstallation of apps [21, 72], understanding users' perceptions of on-device AI security solutions using privacy-sensitive data is essential to promote widespread adoption of the secure solutions.

To address this gap in understanding user perspectives on the trade-offs between security and usability, we investigate how users perceive on-device AI vishing detection apps by posing the following research questions (RQs):

RQ1 How is **smartphone usage satisfaction** affected by on-device AI vishing detection apps?

RQ2 Which **factors influence users' decision to install or recommend** on-device AI vishing detection apps?

Through an in-lab experiment involving 30 participants who installed a prototype of on-device AI vishing detection apps on their own devices, we found that the discomfort caused by the app was negligible compared with their usual smartphone usage. In addition, through interviews, we identified key factors influencing users' adoption of on-device AI vishing detection apps. While we anticipated that on-device AI would alleviate users' privacy concerns, participants still expressed unease, describing sharing phone conversations with the apps as *"being wiretapped"*. Based on our findings, we encourage integrating on-device to develop privacy-preserving mobile security solutions and discuss strategies to promote the widespread adoption of on-device AI systems using private data, such as phone conversations.

2 Related Work

2.1 User Perceived System Overhead of On-Device AI

On-device AI models typically require substantial computational resources [36], which can potentially compromise user experience [41, 72]. Although several studies [14, 22] have examined the computational overhead of AI models on devices, how users perceive this overhead on their smartphones in real-world scenarios has yet to be explored. Moreover, considering that on-device AI vishing detection apps are intended to function in the background during calls, it is crucial to understand their impact on the smartphone user experience under such an execution context. To this end, we developed a prototype app that activates on-device AI speech recognition and text-based vishing detection models in the background when users are engaging in calls. We also conducted in-lab experiments. We analyzed users' perceived system overhead of the app, focusing on its impact on users' smartphone usage satisfaction on their own devices.

2.2 User Perceptions of Sharing Private Data with AI Systems

Several studies [33, 39, 49] have examined user perceptions about sharing private data with AI-based systems, highlighting that users feel discomfort and privacy concerns when data is collected and shared with external vendors or third parties. These perceptions remain in data sharing for services that user enjoy [47]. For instance, users described personalized shopping recommendations based on data collected by the smart speakers as a feeling of being wiretapped [46]. While users expressed less privacy concerns for sharing sensor data with on-device AI than off-device systems [35], users' perceptions of sharing privacy-sensitive data, such as phone conversations, remain uncertain. Recognizing that users sometimes prioritize perceived benefits over privacy concerns when adopting AI-based systems [50, 74, 87], we aim to identify the key factors influencing their willingness to use and recommend on-device AI vishing detection apps.

3 Methodology

We developed a prototype app featuring on-device AI models for vishing detection and measured its system overhead (§ 3.1). We then conducted a user study consisting of in-lab experiments to assess how these apps affect user satisfaction (RQ1) (§ 3.2.1), followed by semi-structured group interviews to identify key factors influencing users' decisions to install or recommend the apps (RQ2) (§ 3.2.2).

3.1 Prototype

We designed an on-device AI vishing detection app prototype to simulate system overhead raised by the apps in a vishing detection scenario. The prototype app executes two on-device AI models in the background upon calls (Fig. 2 ①): a speech recognition model (Fig. 2 ②), which converts spoken words in a call into text, and a text-based vishing detection model (Fig. 2 ③), which assesses whether the transcribed conversation is a vishing attempt.

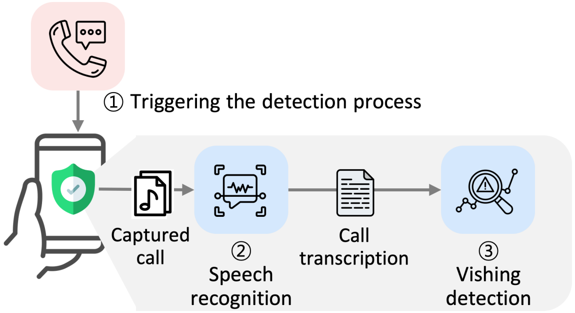


Figure 2: Workflow overview of the prototype app, showcasing the triggered detection process upon receiving a call.

Table 1: System overhead of running our prototype app on a Samsung Galaxy Note 8 (2.6GB of 6GB RAM available).

	Mem. (%)	CPU (%)	Temp. (°C ↑)
On-device AI inferences			
Speech recognition (Wav2Vec 2.0 (base) [10])	3.56	29.99	17.85
Vishing detection (DistilBERT [61])	3.09	43.74	15.05
Prototype app	8.00	40.64	15.85
Common smartphone activities			
Web surfing (Firefox [29])	6.39	3.29	7.17
Watching a video (YouTube [31])	8.64	15.21	8.07
Playing a game (Candy Crush Saga [28])	4.57	13.29	8.67

Note. We observed that all apps use less than 2% of the GPU, indicating that the system overhead primarily stems from CPU usage.

We employed Wav2Vec 2.0 (base) [10] for speech recognition and DistilBERT [61] for vishing detection, chosen for their widely-known efficiency and suitability for deployment on commercial off-the-shelf smartphones [26, 61]. The speech recognition model was trained with 11 million Korean voice samples [2–4, 6, 11, 32, 66]. For testing, 1,000 samples of voice phishing calls were used, resulting in a Word Error Rate (WER) of 14.7%. To train and test the vishing detection model, we utilized 175,732 transcribed Korean real-world vishing calls and messages provided by a Korean government agency responsible for Internet security (KISA) [38], along with non-vishing calls [5]. These data were split into training and testing sets in a 7:3 ratio, achieving a detection accuracy of 94.23%. We utilized PyTorch Mobile [57] to integrate the models into the Android app and replaced call capturing with an example audio clip due to Android restrictions on third-party apps recording calls [27]. We detail the implementation in Appendix A.

Before evaluating users' perceived system overhead of the prototype app, we measured the system overhead in memory usage, CPU utilization, and temperature change on a Samsung Galaxy Note 8 with 2.6GB of available RAM (Table 1).[1] Each of our AI models utilizes up to 3.56% of memory and 43.74% of CPU, with a maximum temperature rise of 17.85°C. When both models are alternately executed, system overhead increases to 8% memory utilization, 40.64% CPU usage, and a temperature rise of 15.85°C. Compared with common smartphone activities (e.g., Firefox for web surfing) [64], AI inferences cause an average increase of 30.04%p in CPU utilization and a 7.88°C rise in temperature. While integrating on-device AI for vishing detection is technically feasible, these overheads emphasize the need for further investigation into user perceptions of the impact on smartphone performance.

3.2 User Study

3.2.1 In-Lab Experiments. To answer RQ1, we designed a controlled environment that simulated real-world smartphone usage conditions and installed the prototype app on participants' own smartphones. Before the in-lab experiment, we explained the study procedure and what the prototype does (e.g., transcribing phone conversations) to our participants. During the experiment, we provided the top three smartphone activities (i.e., web surfing & chatting, watching a video, and playing a game) [64] to mirror real-world smartphone usage. Participants were instructed to perform each activity for 20 minutes and answer a 10-minute call, the average length of Korean vishing calls [38], during each activity. Between activities, there were 3-minute breaks during which participants were asked to leave their smartphones on the desk to cool down. Appendix B includes the full instruction protocol for the experiment.

The 10-minute calls activated the on-device AI vishing detection models in the prototype. By running the models during the calls, we provided participants with a comparable level of system overhead that may occur in real-world on-device AI vishing detection scenarios. To avoid the ethical concerns of exposing participants to real vishing content, we played an audio from the KSponSpeech dataset [11], featuring general open-domain dialogues by native Korean speakers, as a benign alternative throughout the phone call.

For each activity, we conducted two Ecological Momentary Assessment (EMA) surveys: a call survey (requested immediately after each call) and a non-call survey (requested during non-calling periods) to assess changes in usage satisfaction due to perceived system overhead. The following questions were asked on a 5-point Likert scale (1: Very dissatisfied, 5: Very satisfied):

1. **Overheating**: How satisfied are you with *smartphone overheating* at the moment?
2. **App freezes**: How satisfied are you with *app freezes* at the moment?
3. **App crashes**: How satisfied are you with *app crashes* at the moment?
4. **Overall satisfaction**: How satisfied are you with your overall smartphone usage at the moment?

The participants were distributed into the control group (without on-device AI) and the experimental group (with on-device AI). The two groups had the same experimental setting and environment, except the app installed for the control group did not run any on-device AI models during calls.

[1] We chose Samsung Galaxy Note 8, released in 2017, as a representative low-end device to ensure our app's compatibility with a wide range of smartphones rather than focusing on recent high-performance smartphone models.

3.2.2 Interview. We conducted semi-structured group interviews to observe key factors influencing users' decision to install or recommend on-device AI vishing detection apps (RQ2). To begin, participants were asked to compare their general smartphone usage satisfaction and their experiences during the in-lab experiment. We provided a narrative and detailed explanation of on-device AI vishing detection apps, including AI models' roles and on-device processing. Following this introduction, participants were asked about their perceptions of these apps, particularly whether they would consider installing or recommending them to others. All interviews were recorded and transcribed with participant consent, and we report the interview protocol in Appendix C.

3.3 Participants

We recruited 30 participants (aged 19-53, mean=31.8 years; 16 identified as male and 14 as female) through advertisement posts on a popular South Korean online local community platform [19]. Participants were required to sign consent forms stating they agreed to disclose their data. The compensation for each participant was approximately USD 30 for two hours of user study, including the in-lab experiment and the interview. To be eligible for the study, participants had to (1) be over 18 years old and less than 65 years old, (2) use their smartphone daily for web surfing, chatting, playing games, and watching YouTube, and (3) use an Android smartphone. Among study applicants, we selected participants based on the goal of a wide range of ages and diverse smartphone devices. We report detailed demographic and device information in Appendix D.

Participants were equally divided between the control group (aged 19–52, mean=31.9 years; 8 male and 7 female) and the experimental group (aged 19-53, mean=31.7; 8 male and 7 female), ensuring a balance in terms of gender, age, and device types. To keep participants unaware of their group assignment, we installed the prototype app for both groups; however, only the experimental group was configured to run the AI models in the app.

3.4 Data Analysis

To assess the impact of on-device AI vishing detection models on participants' smartphone usage satisfaction, we calculated the difference in satisfaction scores between the call survey and the non-call survey. For the control group, this difference reflects the impact of the call itself (without AI running), while for the experimental group, it captures the combined effect of the call and the execution of the AI models.

We conducted inductive thematic analysis [13] on the interview transcription. The first and third authors independently reviewed all transcripts, developed initial codes for the first two interviews, and discussed emerging themes to reconcile discrepancies. The iterative process continued, analyzing two interviews at a time and updating the codebook periodically. After coding all interviews, the authors reviewed them for consistency and organized the themes and sub-themes into a finalized codebook through comprehensive discussions. The final codebook (Appendix E) consisted of six high-level codes (e.g., usage satisfaction, privacy concerns, etc.) with thirteen codes.

3.5 Ethical Considerations

Our IRB-approved study prioritized ethical research practices and participant privacy by pseudo-anonymizing participants with unique nicknames and informing them of their right to refuse sensitive questions, except for providing their phone number as a key identifier. Considering the serious nature of vishing in South Korea, we avoided exposing participants to deceptive vishing attempts. Instead, we focused on examining changes in smartphone usage satisfaction due to the perceived system overhead of on-device AI models for vishing detection, aligning with our research objectives.

4 Results

4.1 Smartphone Usage Satisfaction with On-Device AI Vishing Detection Models

To evaluate user-perceived impacts of on-device AI vishing detection models on smartphone usage satisfaction (RQ1), we analyzed call and non-call survey responses. Fig. 3 illustrates the distribution of the difference in usage satisfaction between the two surveys: grey represents no change, blue represents a decrease in satisfaction, and orange represents an increase in satisfaction in the call survey compared with the non-call survey.

Overall, most participants in both groups reported no change in satisfaction, even though the prototype app for the experimental group ran on-device AI vishing detection models. Regarding smartphone overheating, a few participants in both groups reported decreased usage satisfaction; however, more than half of the experimental group participants experienced a decrease while performing gaming activities. For app freezes, both groups exhibited similar responses, with a small number of participants reporting decreased satisfaction. Nearly all participants in both groups reported no change in satisfaction related to app crashes. Finally, when evaluating overall satisfaction, very few participants in both groups reported decreased satisfaction. Detailed statistics are in Appendix F.

In the interviews, seven from the control group and six from the experimental group mentioned problems related to overheating, the app freezes, or app crashes during calls. Both groups reported similar reactions despite the prototype app running on-device AI vishing detection models only in the experimental groups' devices. However, four of the six experimental group participants who reported inconveniences mentioned that the overhead was less noticeable than their usual experiences. P22, an experimental group participant, expressed, "*But honestly, the stuttering and heat were way less than I usually feel.*" In summary, though we observed high measured system overhead, it did not significantly affect the usage satisfaction of participants.

4.2 Key Factors Influencing Users' Adoption of On-Device AI Vishing Detection Apps

In the interviews, most participants confirmed their willingness to use or recommend the apps, suggesting their perceived effectiveness against vishing attacks, while few participants hesitated to use the apps. Probing their reasons for adoption decisions revealed three key influencing factors (RQ2): (i) experience with vishing, (ii) age-based perceived vulnerability to threats, and (iii) privacy concerns.

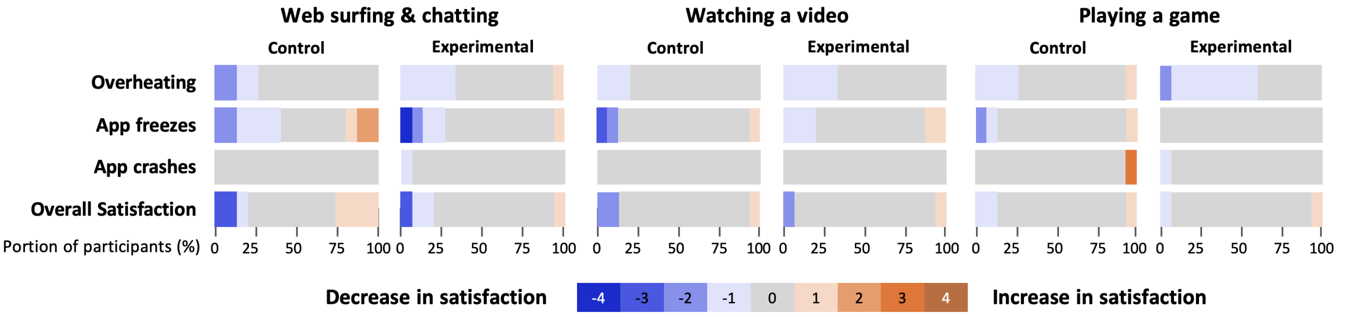


Figure 3: Comparison of usage satisfaction differences between the control and experimental groups.

Experience with vishing. Thirteen participants evaluated on-device AI vishing detection apps as effective countermeasures. Among them, seven participants expressed willingness to install the apps on their phone. Notably, four of them had acquaintances who had experienced vishing incidents, while three of them had nearly fallen victim themselves. P25 mentioned, "*I once received a vishing call and talked for a long time. It was unbelievable how detailed they were with my personal information, and they even knew about my friend's situation in detail, which made me engage with the call. Despite installing an anti-vishing app called 'whowho [30]' [which uses a heuristic approach to detect calls by blacklisting known numbers], it was ineffective in detecting it. If it could filter such [vishing] calls, that would be really helpful.*"

Age-based perceived vulnerability to threats. Seventeen participants said they would recommend on-device AI vishing detection apps to acquaintances who could be vulnerable to attacks, such as the elderly. P4 mentioned, "*I would be willing to install it for my parents. They are more exposed to such scams [vishing] than we are and might impulsively act out of concern for their family.*" However, thirteen participants were confident they would not fall for a vishing attack due to their youth. Five of them explicitly expressed that they would not use the apps for themselves. For example, P6 noted, "*Most young people think they won't be victims. They believe they won't be fooled because they're quite smart. . . . It's not that I have an aversion to it, but I genuinely believe I won't be a victim.*"

Privacy concerns. Despite offering detailed explanations about how on-device AI vishing detection apps work locally, five participants still raised privacy concerns, feeling insecure about sharing data with the app itself, regardless of whether it was processed on-device or off-device. P8 stated, "*I know it only converts phone conversations to text on my phone, but I'm somewhat concerned about privacy, so I don't think I'll use it.*" Furthermore, three of them were emotionally uneasy about sharing phone conversation with the apps, stating "*Converting private phone conversation into text through AI may not pose a problem, but it still feels unpleasant*" (P21). P17 even anthropomorphized the apps, stating "*To be honest, if I exaggerate a bit, it might feel like being wiretapped.*"

5 Discussion

Our study aimed to understand users' perceptions of on-device AI vishing detection apps that utilize privacy-sensitive data (i.e., phone conversations). We found that users' perceived overhead of two on-device AI models used by the prototype app was negligible compared with their daily discomfort. In addition, we identified experience with vishing, age-based perceived vulnerability to threats, and privacy concerns as key factors influencing users' adoption of on-device AI vishing detection apps. Based on our findings, we encourage integrating on-device AI to develop privacy-preserving mobile security solutions and discuss strategies to promote the widespread adoption of on-device AI systems using privacy-sensitive data, such as phone conversations.

5.1 Integrating On-Device AI into Mobile Security

We encourage security researchers to actively adopt AI-based solutions to mobile security. Unlike previous work focusing on high-end devices such as the Samsung Galaxy S2 Plus [35], our study employed a diverse range of devices by having participants use their smartphones. This approach allowed us to evaluate the perceived system overhead of on-device AI across a much broader spectrum of device specifications, and our participants similarly confirmed that the perceived system overhead was comparable to typical smartphone usage. A potential future direction could involve applying our solution to security solutions to combat mobile text-based scams, such as smishing [75, 82] or direct messaging-based scams [81].

Different attacks require different on-device AI solutions, each with a tailored execution scenario of AI models. As we considered the scenario of taking phone calls in evaluating the perceived system overhead for vishing detection, it is essential to account for the distinct context of each attack scenario when assessing user perceptions of such systems. We expect future research on extended mobile attacks (e.g., malware [59] and wireless attacks [12, 70]) to focus on their perceived system overhead specifically in the context of their tailored use scenarios.

5.2 Enhancing Perceived Vulnerability of Vishing

Many participants recognized the effectiveness of on-device AI-based solutions against vishing and were willing to recommend them to vulnerable groups, such as older adults and children. However, participants often perceived themselves as immune to vishing threats, believing they could identify such attacks through intuition. This confidence led them to undervalue the usefulness of these secure solutions for their personal use, although vishing can effectively target individuals of all ages and educational backgrounds [42, 65, 76].

While public campaigns have long aimed to raise awareness of vishing vulnerabilities [25, 45, 63], our findings highlight the continued importance of educating users to adopt secure behaviors. Recent research emphasizes that interactive methods, such as role-playing, can significantly enhance support-seeking behavior against phishing [16]. Building on this approach, further research should explore practical strategies like personalized, context-aware warnings that translate abstract security concepts into tangible, individual risk assessments [53]. For instance, a user who frequently discusses financial matters over the phone could receive a tailored warning, such as, *"You are at high risk of mistaking a vishing call for a regular conversation about financial transfers."*

5.3 Resolving Persistent Privacy Concerns in On-Device AI Systems

Despite the local processing of data by on-device AI vishing detection apps, participants expressed privacy concerns, with some likening the experience to *"being wiretapped"* (P17), reflecting misconceptions similar to those observed with off-device systems [46]. While users prefer sharing sensor data (e.g., accelerometer or gyroscope) with on-device systems over off-device ones [35], our findings highlight privacy concerns about sharing phone conversation data. This may stem from how users prioritize different types of data privacy, with sensitive data like medical records and browsing history often valued more highly than physical activity data [69].

Importantly, we found that incomplete or inaccurate mental models of on-device AI systems can lead to privacy concerns. Notably, some participants indicated their understanding of the system's data processing by mentioning, *"only converts phone conversations to text on my phone"* (P8). However, the persistence of these concerns suggests they may not fully grasp its on-device nature. To address these concerns, on-device AI apps such as AI vishing detection tools should incorporate effective communication mechanisms to enhance user trust and understanding. For example, messages displayed during inference (e.g., *"Your data is securely processed locally on your device"*), can clarify system operations and increase user confidence [71]. In addition, user interfaces illustrating local data processing and automatic data disposal may help users develop a more accurate mental model and avoid misconceptions. AI security app developers should integrate such design elements. However, technical solutions alone may not completely address all issues, highlighting the need for broader efforts.

Effective policy frameworks are also crucial for building user trust in on-device AI apps. Governments should establish clear guidelines and certification programs for apps handling sensitive data like phone conversations. Regular audits will ensure compliance, reinforce credibility, and maintain public trust in these security solutions. Furthermore, governments should promote these certification programs and provide a list of certified on-device AI solutions. This can be achieved through media campaigns, educational initiatives, and collaborations with solution vendors. By strengthening regulatory oversight and fostering transparency, governments can help reduce user mistrust and encourage the adoption of secure on-device AI solutions.

5.4 Limitations

We recruited participants with diverse ages, genders, and occupations to capture a broad range of user perceptions. However, all participants were South Korean, limiting national diversity. Additionally, most used Samsung smartphones with Android 13, reflecting South Korea's 2024 trend, where 90% of Android users prefer Samsung devices [44]. Future research should include participants from diverse nationalities and cultural backgrounds to better account for variations in technological familiarity and privacy concerns.

Due to Android restrictions on third-party apps' call recordings, the prototype app analyzes pre-stored example audio files instead of actual call recordings. Recent advancements, such as a Korean telecommunication company's AI phone conversation summary service using redirected and recorded calls via mVoIP [55], suggest potential methods for overcoming these limitations in future work.

6 Conclusion

We explored users' (mis)perceptions of on-device AI apps for vishing detection. Through in-lab experiments with a prototype app featuring two on-device AI models and interviews with 30 participants, we found that the impact of on-device AI models for vishing detection on smartphone usage satisfaction was insignificant. Additionally, we identified experience with vishing, age-based perceived vulnerability to threats, and privacy concerns as key factors in adopting the apps. Notably, some participants showed persistent privacy concerns despite the system's on-device processing to ensure data security. Our findings illuminate the importance of developing privacy-preserving on-device AI solutions while enhancing users' understanding to promote broader adoption.

Future research includes designing and developing a fully functional on-device AI vishing detection app with effective interventions, including warning screens and real-time alerts. Incorporating explainable AI (XAI) features that provide users with clear explanations of why a call is detected as vishing would enhance transparency and foster trust. Field studies in real-world scenarios will further help to evaluate the app's effectiveness, provide deeper insights into users' mental models, and guide refinements to ensure interventions are both informative and sensitive to privacy concerns, such as the feeling of being wiretapped.

Acknowledgments

This work is funded by the National Research Foundation of Korea (NRF), funded by the Ministry of Science and ICT (MSIT) under grant RS-2024-00464269 and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00337007). ※ MSIT: Ministry of Science and ICT.

References

[1] Korean National Police Agent. 2024. https://www.data.go.kr/data/15063815/fileData.do. Accessed: March 3, 2025.

[2] AIHub. 2019. KoreanSpeech dataset. https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=123. Accessed: March 3, 2025.

[3] AIHub. 2021. Counseling call dataset. https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=100. Accessed: March 3, 2025.

[4] AIHub. 2022. Welfare call center counseling dataset. https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&dataSetSn=470. Accessed: March 3, 2025.

[5] AIHub. 2023. Korean complain call dataset. https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&dataSetSn=98.

[6] AIHub. 2023. Low-quality call voice recognition dataset. https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=data&dataSetSn=571. Accessed: March 3, 2025.

[7] Anti Scam. 2024. Anti Scam. https://www.antiscam.co.kr. Accessed: March 3, 2025.

[8] Farhanim Mohamad Asri and Tengku Elena Tengku Mahamad. 2023. Anatomy of Phone Scams: Victims' Recall on the Communication Phrases used by Phone Scammers. In *International Conference on Communication and Media 2022 (i-COME 2022)*. Atlantis Press, 498–509.

[9] Avast. 2019. Voice fraud scams company out of $243,000. https://blog.avast.com/deepfake-voice-fraud-causes-243k-scam. Accessed: March 3, 2025.

[10] Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. 2020. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems* 33 (2020), 12449–12460.

[11] Jeong-Uk Bang, Seung Yun, Seung-Hi Kim, Mu-Yeol Choi, Min-Kyu Lee, Yeo-Jeong Kim, Dong-Hyun Kim, Jun Park, Young-Jik Lee, and Sang-Hun Kim. 2020. Ksponspeech: Korean spontaneous speech corpus for automatic speech recognition. *Applied Sciences* 10, 19 (2020), 6936.

[12] Arup Barua, Md Abdullah Al Alamin, Md Shohrab Hossain, and Ekram Hossain. 2022. Security and privacy threats for bluetooth low energy in iot and wearable devices: A comprehensive survey. *IEEE Open Journal of the Communications Society* 3 (2022), 251–281.

[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[14] Han Cai, Ji Lin, Yujun Lin, Zhijian Liu, Haotian Tang, Hanrui Wang, Ligeng Zhu, and Song Han. 2022. Enable deep learning on mobile devices: Methods, systems, and applications. *ACM Transactions on Design Automation of Electronic Systems (TODAES)* 27, 3 (2022), 1–50.

[15] Rajasekhar Chaganti, Bharat Bhushan, Anand Nayyar, and Azrour Mourade. 2021. Recent trends in social engineering scams and case study of gift card scam. *arXiv preprint arXiv:2110.06487* (2021).

[16] Xiaowei Chen, Margault Sacré, Gabriele Lenzini, Samuel Greiff, Verena Distler, and Anastasia Sergeeva. 2024. The Effects of Group Discussion and Role-playing Training on Self-efficacy, Support-seeking, and Reporting Phishing Emails: Evidence from a Mixed-design Experiment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–21.

[17] Federal Trade Commission. 2023. New FTC data show consumers reported losing nearly $8.8 billion to scams in 2022. https://www.ftc.gov/news-events/news/press-releases/2023/02/new-ftc-data-show-consumers-reported-losing-nearly-88-billion-scams-2022. Accessed: March 3, 2025.

[18] Federal Trade Commission. 2024. FTC Imposter Scams. https://www.ftc.gov/office-inspector-general/ftc-imposter-scams. Accessed: March 3, 2025.

[19] Karrot Corp. 2024. Karrot. https://karrotmarket.com/?in=manhattan-7426.

[20] China Daily. 2023. Shanghai police warn of FaceTime scams. https://www.chinadaily.com.cn/a/202307/09/WS64aaa018a310bf8a75d6e12b.html. Accessed: March 3, 2025.

[21] Kenan Degirmenci. 2020. Mobile users' information privacy concerns and the role of app permission requests. *International Journal of Information Management* 50 (2020), 261–272.

[22] Yunbin Deng. 2019. Deep learning on mobile devices: a review. *Mobile Multimedia/Image Processing, Security, and Applications 2019* 10993 (2019), 52–66.

[23] Stacy Cowley Emily Flitter. 2023. Voice Deepfakes Are Coming for Your Bank Balance. https://www.nytimes.com/2023/08/30/business/voice-deepfakes-bank-scams.html. Accessed: March 3, 2025.

[24] Financial Supervisory Service. 2024. Voice Phishing Guidance. https://www.fss.or.kr/fss/main/sub1voice.do?menuNo=200012. Accessed: March 3, 2025.

[25] ftd2022who 2022. Who experiences scams? A story for all ages. https://www.ftc.gov/news-events/data-visualizations/data-spotlight/2022/12/who-experiences-scams-story-all-ages. Accessed: March 3, 2025.

[26] Santosh Gondi. 2022. Wav2vec2. 0 on the edge: Performance evaluation. *arXiv preprint arXiv:2202.05993* (2022).

[27] Google Phone App Help. 2024. Use the Phone app to record calls. https://support.google.com/phoneapp/answer/9803950?hl=en&sjid=5970300710915922370-AP. Accessed: March 3, 2025.

[28] Google Play Store. 2024. Candycrush. https://play.google.com/store/apps/details?id=com.king.candycrushsago. Accessed: March 3, 2025.

[29] Google Play Store. 2024. Firefox. https://play.google.com/store/apps/details?id=org.mozilla.firefox. Accessed: March 3, 2025.

[30] Google Play Store. 2024. Who Who. https://play.google.com/store/apps/details?id=com.ktcs.whowho&hl=en&gl=US. Accessed: March 3, 2025.

[31] Google Play Store. 2024. YouTube. https://play.google.com/store/apps/details?id=com.google.android.youtube. Accessed: March 3, 2025.

[32] Jung-Woo Ha, Kihyun Nam, Jingu Kang, Sang-Woo Lee, Sohee Yang, Hyunhoon Jung, Eunmi Kim, Hyeji Kim, Soojin Kim, Hyun Ah Kim, et al. 2020. Clovacall: Korean goal-oriented dialog speech corpus for automatic speech recognition of contact centers. *arXiv preprint arXiv:2004.09367* (2020).

[33] Muhammad Haris, Hamed Haddadi, and Pan Hui. 2014. Privacy leakage in mobile computing: Tools, methods, and characteristics. *arXiv preprint arXiv:1410.4978* (2014).

[34] Sumair Ijaz Hashmi, Niklas George, Eimaan Saqib, Fatima Ali, Nawaal Siddique, Shafay Kashif, Shahzaib Ali, Nida Ul Habib Bajwa, and Mobin Javed. 2023. Training Users to Recognize Persuasion Techniques in Vishing Calls. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–8.

[35] Jun Ho Huh, Sungsu Kwag, Iljoo Kim, Alexandr Popov, Younghan Park, Geumhwan Cho, Juwon Lee, Hyoungshick Kim, and Choong-Hoon Lee. 2023. On the Long-Term Effects of Continuous Keystroke Authentication: Keeping User Frustration Low through Behavior Adaptation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (2023), 1–32.

[36] Andrey Ignatov, Radu Timofte, William Chou, Ke Wang, Max Wu, Tim Hartley, and Luc Van Gool. 2018. Ai benchmark: Running deep neural networks on android smartphones. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 0–0.

[37] This Week in Asia. 2021. Overseas chinese in japan warned on phone scams demanding bank transfers. https://www.scmp.com/week-asia/lifestyle-culture/article/3116336/overseas-chinese-japan-warned-phone-scams-demanding. Accessed: March 3, 2025.

[38] Korea Internet and Security Agency. 2017. Korea Internet and Security Agency. https://www.kisa.or.kr/. Accessed: March 3, 2025.

[39] Timo Jakobi, Corinna Ogonowski, Nico Castelli, Gunnar Stevens, and Volker Wulf. 2017. The catch (es) with smart home: Experiences of a living lab field study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1620–1633.

[40] Keith S Jones, Miriam E Armstrong, McKenna K Tornblad, and Akbar Siami Namin. 2021. How social engineers use persuasion principles during vishing attacks. *Information & Computer Security* 29, 2 (2021), 314–331.

[41] Soowon Kang, Hyeonwoo Choi, Sooyoung Park, Chunjong Park, Jemin Lee, Uichin Lee, and Sung-Ju Lee. 2019. Fire in your hands: Understanding thermal behavior of smartphones. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.

[42] Joongyum Kim, Jihwan Kim, Seongil Wi, Yongdae Kim, and Sooel Son. 2022. HearMeOut: detecting voice phishing activities in Android. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*. 422–435.

[43] Kyusik Kim. 2024. LG U+ has entered the competition for AI secretaries. https://www.mk.co.kr/en/it/11162530. Accessed: March 3, 2025.

[44] Gallup Korea. 2024. Smartphone statistics 2012-2024. https://www.gallup.co.kr/gallupdb/reportContent.asp?seqNo=1497. Accessed: March 3, 2025.

[45] Jim Kreidler. 2023. Scam proof the young people in your life. https://consumer.ftc.gov/consumer-alerts/2023/05/scam-proof-young-people-your-life. Accessed: March 3, 2025.

[46] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–31.

[47] Sunok Lee, Minji Cho, and Sangsu Lee. 2020. What if conversational agents became invisible? comparing users' mental models according to physical entity of ai speaker. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–24.

[48] Federico Maggi. 2010. Are the con artists back? a preliminary analysis of modern phone frauds. In *2010 10th IEEE International Conference on Computer and Information Technology*. IEEE, 824–831.

[49] Karola Marky, Sarah Prange, Max Mühlhäuser, and Florian Alt. 2021. Roles matter! Understanding differences in the privacy mental models of smart home visitors and residents. In *Proceedings of the 20th International Conference on Mobile and Ubiquitous Multimedia*. 108–122.

[50] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. 2017. Privacy expectations and preferences in an {IoT} world. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. 399–412.

[51] BBC News. 2021. Nearly 45 million received scam calls in three months, Ofcom says. https://www.bbc.com/news/technology-58982233. Accessed: March 3, 2025.

[52] Sharbani Pandit, Krishanu Sarker, Roberto Perdisci, Mustaque Ahamad, and Diyi Yang. 2023. Combating robocalls with phone virtual assistant mediated interaction. In *USENIX Security*.

[53] Eyal Peer, Serge Egelman, Marian Harbach, Nathan Malkin, Arunesh Mathur, and Alisa Frik. 2020. Nudge me right: Personalizing online security nudges to people's decision-making styles. *Computers in Human Behavior* 109 (2020), 106347.

[54] PhishingEyes. 2024. Phishing Eyes. https://www.phishingeyes.com. Accessed: March 3, 2025.

[55] Google Playstore. 2024. ADOT. https://play.google.com/store/apps/details?id=com.skt.nugu.apollo&hl=en. Accessed: March 3, 2025.

[56] Alvaro Pulg. 2023. Scammers are impersonating FTC Inspector General Andrew Katsaros. https://consumer.ftc.gov/consumer-alerts/2023/10/scammers-are-impersonating-ftc-inspector-general-andrew-katsaros. Accessed: March 3, 2025.

[57] PyTorch. 2019. PyTorch Mobile. https://pytorch.org/mobile/home/. Accessed: March 3, 2025.

[58] Lubna Razaq, Tallal Ahmad, Samia Ibtasam, Umer Ramzan, and Shrirang Mare. 2021. "We Even Borrowed Money From Our Neighbor" Understanding Mobile-based Frauds Through Victims' Experiences. *Proceedings of the ACM on human-computer interaction* 5, CSCW1 (2021), 1–30.

[59] Asma Razgallah, Raphaël Khoury, Sylvain Hallé, and Kobra Khanmohammadi. 2021. A survey of malware detection in Android apps: Recommendations and perspectives for future research. *Computer Science Review* 39 (2021), 100358.

[60] Merve Sahin, Aurélien Francillon, Payas Gupta, and Mustaque Ahamad. 2017. Sok: Fraud in telephony networks. In *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 235–250.

[61] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108* (2019).

[62] Scamwatch. 2024. Scamwatch. https://www.scamwatch.gov.au/. Accessed: March 3, 2025.

[63] scamwatch2020genz 2020. Gen Z the fastest growing victims of scams. https://www.scamwatch.gov.au/news-alerts/gen-z-the-fastest-growing-victims-of-scams. Accessed: March 3, 2025.

[64] Common sense. 2021. The common Sense Census: Media Use by Tweens and Teens.

[65] Oh Seok-min. 2018. Youths more prone to voice phishing scams than elderly: data. https://en.yna.co.kr/view/AEN20181030006600320. Accessed: March 3, 2025.

[66] Korean Financial Supervisory Service. 2008. Korean Financial Supervisory Service. https://www.fss.or.kr/. Accessed: March 3, 2025.

[67] Sanket Shukla, PD Sai Manoj, Gaurav Kolhe, and Setareh Rafatirad. 2021. On-device malware detection using performance-aware and robust collaborative learning. In *2021 58th ACM/IEEE Design Automation Conference (DAC)*. IEEE, 967–972.

[68] Stu Sjouwerman. 2024. Deepfake Phishing: The Dangerous New Face Of Cybercrime. https://www.forbes.com/sites/forbestechcouncil/2024/01/23/deepfake-phishing-the-dangerous-new-face-of-cybercrime/?sh=4fd0b3564aed. Accessed: March 3, 2025.

[69] Anya Skatova, Rebecca McDonald, Sinong Ma, and Carsten Maple. 2023. Unpacking privacy: Valuation of personal data protection. *Plos one* 18, 5 (2023), e0284581.

[70] Raphael Spreitzer, Veelasha Moonsamy, Thomas Korak, and Stefan Mangard. 2017. Systematic classification of side-channel attacks: A case study for mobile devices. *IEEE communications surveys & tutorials* 20, 1 (2017), 465–488.

[71] Christian Stransky, Dominik Wermke, Johanna Schrader, Nicolas Huaman, Yasemin Acar, Anna Lena Fehlhaber, Miranda Wei, Blase Ur, and Sascha Fahl. 2021. On the Limited Impact of Visualizing Encryption: Perceptions of E2E Messaging Security. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*. USENIX Association, 437–454. https://www.usenix.org/conference/soups2021/presentation/stransky

[72] Hyewon Suh, Nina Shahriaree, Eric B Hekler, and Julie A Kientz. 2016. Developing and validating the user burden scale: A tool for assessing user burden in computing systems. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 3988–3999.

[73] Jun sung Lee. 2024. SKT Sets New Standard for AI Calls with 'A.Phone'. https://www.koreaittimes.com/news/articleView.html?idxno=135224. Accessed: March 3, 2025.

[74] Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. 2019. Investigating users' preferences and expectations for always-listening voice assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–23.

[75] Sarah Tabassum, Cori Faklaris, and Heather Richter Lipford. 2024. What Drives {SMiShing} Susceptibility? A {US}. Interview Study of How and Why Mobile Phone Users Judge Text Messages to be Real or Fake. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*. 393–411.

[76] Truecaller. 2022. 2022 U.S. Spam & Scam Rport. https://www.truecaller.com/blog/insights/truecaller-insights-2022-us-spam-scam-report. Accessed: March 3, 2025.

[77] Truecaller. 2023. 2023 U.S. Spam & Scam Rport. https://www.truecaller.com/blog/insights/the-true-cost-of-spam-and-scam-calls-in-america.

[78] Truecaller. 2024. Truecaller. https://www.truecaller.com/. Accessed: March 3, 2025.

[79] Huahong Tu, Adam Doupé, Ziming Zhao, and Gail-Joon Ahn. 2016. Sok: Everyone hates robocalls: A survey of techniques against telephone spam. In *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, 320–338.

[80] Huahong Tu, Adam Doupé, Ziming Zhao, and Gail-Joon Ahn. 2019. Users really do answer telephone scams. In *28th USENIX Security Symposium (USENIX Security 19)*. 1327–1340.

[81] Raj Vardhan, Alok Chandrawal, Phakpoom Chinprutthiwong, Yangyong Zhang, and Guofei Gu. 2023. # DM-Me: Susceptibility to Direct Messaging-Based Scams. In *Proceedings of the 2023 ACM Asia Conference on Computer and Communications Security*. 494–508.

[82] Chenkai Wang, Zhuofan Jia, Hadjer Benkraouda, Cody Zevnik, Nicholas Heuermann, Roopa Foulger, Jonathan A Handler, and Gang Wang. 2024. VeriSMS: A Message Verification System for Inclusive Patient Outreach against Phishing Attacks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[83] Whoscall. 2024. Whoscall. https://www.whoscall.com/en. Accessed: March 3, 2025.

[84] Heng Xu, Sumeet Gupta, Mary Beth Rosson, and John Carroll. 2012. Measuring mobile users' concerns for information privacy. *International Conference on Information Systems, ICIS 2012* 3 (01 2012), 2278–2293.

[85] Xiangyu Xu, Jiadi Yu, Yingying Chen, Qin Hua, Yanmin Zhu, Yi-Chao Chen, and Minglu Li. 2020. TouchPass: Towards behavior-irrelevant on-touch user authentication on smartphones leveraging vibrations. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–13.

[86] YouMail. 2024. YouMail. https://www.youmail.com/. Accessed: March 3, 2025.

[87] Bo Zhang, Na Wang, and Hongxia Jin. 2014. Privacy concerns in online recommender systems: influences of control and user data input. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. 159–173.

A System overview of the prototype app

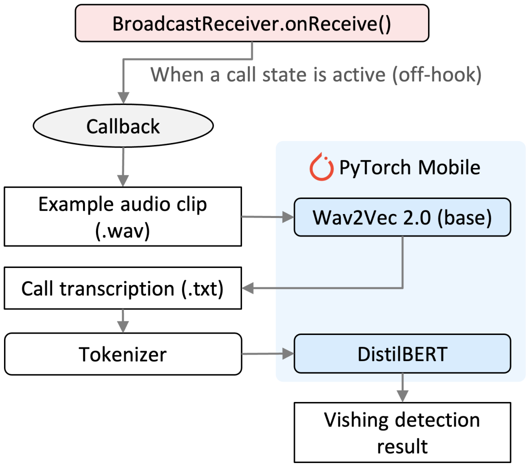

BroadcastReceiver.onReceive()
When a call state is active (off-hook)
Callback
PyTorch Mobile
Example audio clip (.wav)
Wav2Vec 2.0 (base)
Call transcription (.txt)
Tokenizer
DistilBERT
Vishing detection result

Figure 4: System overview of the prototype app.

Figure 4 shows the system overview of our Anroid prototype app for on-device AI vishing detection apps. The rounded rectangles represent the components we implemented, the rectangles represent the outputs of each component, and the oval represents the Android component. We utilized the Android `BroadcastReceiver.onReceive` callback function to trigger on-device AI vishing detection upon receiving a call. The function was implemented to run in the background without disrupting users during the call. The two AI models we trained for speech recognition and text-based vishing detection were integrated into the prototype app using PyTorch Mobile [57]. The trained models were first converted into serialized formats using TorchScript, ensuring compatibility with the Android Java compile environment. For analyzing the call data, every 15 seconds, the audio stream, sampled at 16kHz, was transformed into tensor arrays over a 15-second window. The output from the speech recognition model, consisting of transcribed text, was used as input for the vishing detection model. To ensure synchronized execution, each model function was linked through callback functions. To process Korean text inputs and facilitate data transfer between the models, we utilized a mobile Korean tokenizer that encodes Korean words into 8,002 different tokens.

Given that Android restricts third-party apps from capturing live call audio to protect user privacy and security [27], we used an example audio clip from KSponSpeech dataset [11] as input for the vishing detection instead of recorded call audio. This approach allows us to simulate the system overhead of on-device AI vishing detection apps while preserving participants' privacy and maintaining device integrity without rooting participants' smartphones. In addition, to confirm that the AI models were active during the experiment, we logged the inference results of the example audio clip to Firebase in real time.

B User Study Instruction

Hello. Thank you once again for participating in the study. Here are some details about the experiment. Feel free to ask if you have any questions.

First, you will use a nickname instead of your name during the study. Your assigned nickname is {nickname}. This nickname will be used 1) when entering the KakaoTalk 1:1 open chat room and 2) when participating in surveys (Google Forms) during the experiment.

1. **Study goal**
   This study aims to understand the impact of an AI app on user experience. As previously informed, we do not collect or store any of your private data. Throughout the experiment, our app will only convert preloaded audio files into text and not use or store any of your data.
2. **Study procedure**
   The study is divided into two main parts over a total of two hours: one hour for a 'Smartphone Usage Task and Survey' and another hour for a 'Group Interview.' The first part consists of three sessions. Each session will take about 20 minutes, and you will get one phone call and two survey notifications during each session. There will be a 3-minute break between sessions, and you are not allowed to use your smartphone during the break.

2-1. Smartphone Usage Task and Survey
We prepared three smartphone usage scenarios to explore how our app affects your daily smartphone usage. Detailed instructions for each session will be provided right before it.
(1) **Web surfing & chatting**: Plan a 4-day trip to Jeju Island by searching on Firefox and sharing the plan with staff via KaKaoTalk.
(2) **Watching a video**: Watch a given YouTube video clip in full screen and video quality 1080p.
(3) **Playing a game**: Play a game, Candy Crush Saga, for 20 minutes

2-2. Phone call
During the sessions, please pick up the call from us once in 20 minutes. The call will play a meaningless, random Korean conversion audio clip. You don't need to answer anything and can continue doing the given task while listening to the call. For instance, you can play the game while listening to the call.

2-3. Survey
During the sessions, the app will send you a survey notification twice in 20 minutes. Please click the link and submit the survey (Google Form).

3. Preparation before the experiment
Following are some key points you'll need to check. If you're having trouble with any settings, please feel free to ask staff for help.

- Close all background apps (current app list - close all).
- Disable the 'Priority mode' of the game booster if it's enabled on your phone.
- If you have set up blocking for notifications and calls while watching YouTube, make sure to disable it.

4. Guidelines for the experiment
Please keep the following in mind during the experiment:

- The call will be from {number for experiment}. If you accidentally reject the call, please inform us right away.
- Reject any calls that are not from us. If you receive an unavoidable call, please let us know.
- It's important not to do any other activities outside the given tasks (web surfing, chatting, watching YouTube, and playing the game). Please avoid using any apps unrelated to the current session, and do not use permitted apps for purposes other than those specified above. If you encounter any unavoidable issues during the experiment, please let us know immediately.

If you have any other questions, please feel free to ask.

C Interview Protocol

This is an English-translated summary of the key questions we aimed to address. In the semi-structured interview, we had the opportunity to delve deeper and investigate topics beyond these specific questions.

Warm Up

Hello, Thank you once again for participating in this user study. As previously informed, the interview will last up to one hour. Please feel free to respond in any manner you're comfortable with. If there are any questions during the interview that you don't want to answer, you are welcome to decline to respond. As agreed in the consent form, our interview will be audio-recorded for later analysis, and the data will not be used for any purpose other than research. Do you have any questions before we begin recording?

This experiment was conducted to investigate the impact of our prototype app on usability. Before the interview, I will explain the main purpose of our research.

Introducing Vishing

We are currently engaged in research aimed at proactively preventing vishing. This type of financial fraud often involves criminals posing as representatives of public institutions such as banks or the police, coercing victims into transferring money. The National Police Agency reported that in 2022, there were over 20,000 incidents of vishing, resulting in damages surpassing 500 billion won.

In these vishing calls, scammers typically impersonate authoritative bodies or leverage leaked personal data to gain the victim's trust. They frequently resort to tactics like intimidation or creating a false sense of urgency, making it challenging for victims to recognize the deceit. Notably, a recent tactic involves persuading victims to download counterfeit applications that closely mimic genuine banking apps, through which personal information is extracted.

Debriefing Purpose of User Study

To effectively avoid falling for vishing, it is essential to identify and disengage from the calling situation quickly. In pursuit of this goal, we are exploring the development of an application designed to alert users of suspected vishing attempts. This app utilizes AI technology to assess whether a phone conversation could be a vishing attempt. Before the actual development of a vishing detection app, today's experiment was designed to explore whether such an AI-powered application causes any inconvenience in your smartphone usage.

As previously explained, the prototype app installed on your smartphones for this experiment is a basic AI app. It functions similarly to a language translator, but instead of converting Korean to English, it transforms spoken words into written text. Throughout the experiment, the app's sole task was transcribing pre-loaded voice files into text. It did not access or store any of your personal data. You can be assured that the app's functionality was strictly confined to this transcription process.

Do you have any questions before the interview?

Typical Smartphone Usages

First, I'd like to ask in detail about any inconveniences you usually experience while using your smartphone. Remember the questionnaire you filled out before the experiment started? It asked about the discomfort you experience while web surfing, watching YouTube, and playing games on your phone, focusing on issues like overheating, app freeze, and app crash. I'd like to hear more about your experiences related to these issues.

For each task,

(1) Please share your experience of discomforts like overheating, app freeze, or app crash.
(2) If you have experienced these discomforts without finding them inconvenient, could you please explain why you perceive them in this way?

User Experience during the Experiment

I'm curious about how today's experiment compared to your usual smartphone usage experience.

For each task,

(1) Did you notice any significant differences in overheating, app freeze, or app crash? Were there any discomforts related to these aspects during the experiment?

Vishing Detection Apps

It seems we can wrap up the questions about your usage experience. Now, let's discuss the app we're developing, which detects and alerts users about vishing. This vishing alert app operates only during phone calls and utilizes two AI functions to determine whether a call is a phishing attempt. The first function converts the sounds of the call into text like a translator converts Korean to English. The second function analyzes this text to identify whether it is vishing. These AI features operate locally on the smartphone rather than sending data to larger computers in a data center or cloud for processing.

Do you have any questions about this app?

Now, I'd like you to imagine receiving a call from a vishing criminal.

(1) How would you feel if an AI app automatically activated during such a call, detecting and warning you of the vishing attempt?
(2) Would you be interested in downloading and using such an app?
(3) Would you recommend it to family or friends?
    (For those who reported significant discomfort during the experiment)
    (a) Considering your experience with smartphone overheating, app freeze, and app crash during today's experiment, would you still be interested in using a vishing warning app?

Closing

Thank you sincerely for answering all the questions diligently. We have asked everything we wanted to know. Before we conclude the interview, is there anything else you would like to share with us or any responses you feel you didn't get a chance to express?

With that, we will conclude today's interview. This document is a personal information consent form to pay the participation fee. Please fill it out, and you can delete the installed experimental app. When you submit the form, we'll help confirm that the app has been deleted. We will also send you a text message regarding the payment of the participation fee.

Once again, thank you very much for participating.

D Participants Information

Table 2: Participants demographic, vishing experience, and device information. The number next to the group indicates subgroups for the group interview. vishing experience represents who has experienced vishing (Direct), who knows close acquaintance with vishing experience (Indirect), and who has heard of vishing (Aware of).

P	Group	Age	Gender	Education	Occupation	Vishing experience			Device	Available RAM (GB)	Android version
						Direct	Indirect	Aware of			
1	cont 1	19	M	Undergraduate	Student			✓	Galaxy S8	1.1	9
2	cont 1	26	M	Bachelor's	Student			✓	Galaxy S23	2.3	13
3	cont 1	35	M	Bachelor's	Freelancer/Professional		✓		Galaxy S21	1.2	9
4	cont 1	38	F	Bachelor's	Administrative/Technical		✓	✓	Galaxy Note 9	2.1	10
5	cont 2	24	M	Bachelor's	Student		✓	✓	Galaxy Note 8	2.2	9
6	cont 2	33	F	Bachelor's	Administrative/Technical			✓	Galaxy Flip 4	3.4	13
7	cont 2	49	F	Bachelor's	Freelancer/Professional			✓	Galaxy Note 8	2.2	9
8	cont 3	19	F	Undergraduate	Student			✓	Galaxy S10	2.6	12
9	cont 3	19	M	Undergraduate	Student			✓	Galaxy A53 5G	1.5	13
10	cont 3	20	M	Not to disclose	Not to disclose	-	-	-	Galaxy A53	1.4	13
11	cont 3	26	M	Bachelor's	Administrative/Technical		✓		Galaxy Note 20	2.7	13
12	cont 3	33	M	Ph.D.	Administrative/Technical			✓	Galaxy S9	1.2	9
13	cont 3	40	F	Bachelor's	Homemaker			✓	Galaxy Note 4	-	6
14	cont 3	46	F	Bachelor's	Homemaker			✓	Galaxy S22 Ultra	5.3	13
15	cont 3	52	F	Bachelor's	Homemaker		✓		Galaxy S22	2.1	13
16	exp 1	22	M	Undergraduate	Student	✓			Galaxy S21 Plus	2.1	13
17	exp 1	23	M	Undergraduate	Student			✓	Galaxy Note 10 Plus	6	12
18	exp 1	24	M	Undergraduate	Student		✓	✓	Galaxy S22	2.3	13
19	exp 1	34	F	Bachelor's	Freelancer/Professional		✓	✓	Galaxy Note 20 Ultra	2.6	13
20	exp 1	44	M	Master's	Others		✓	✓	Galaxy Flip 4	2.5	13
21	exp 2	19	M	Undergraduate	Student			✓	Galaxy A32	0.572	13
22	exp 2	21	M	Undergraduate	Student			✓	Galaxy S21	2.5	13
23	exp 2	24	M	Not to disclose	Not to disclose	-	-	-	Galaxy S22	2.3	13
24	exp 2	30	F	Master's	Administrative/Technical			✓	Galaxy S20	3.8	13
25	exp 2	35	F	Master's	Others	✓			Galaxy S20 FE	1.4	13
26	exp 3	22	M	Undergraduate	Student	✓			Galaxy S22 Ultra	4.3	13
27	exp 3	41	F	Bachelor's	Freelancer/Professional		✓		Galaxy S22	2.6	13
28	exp 3	42	F	Bachelor's	Freelancer/Professional		✓	✓	Galaxy S20 Plus	2.5	13
29	exp 3	42	F	Bachelor's	Homemaker	✓	✓		Galaxy S22 Ultra	4.6	13
30	exp 3	53	F	Bachelor's	Administrative/Technical	✓	✓		LG Q520N	0.689	12

Note. P10 and P23 did not answer the optional questions, and P13's device did not provide information on available RAM capacity.

E Codebook

High-level Code	Subcodes	Illustrative Quotations
Adoption of on-device AI voice phishing detection apps	Install the apps on children's phones	*"And these days, even young children are using smartphones, so I would definitely talk to them and likely make a lot of recommendations to them."* (P13)
	Install/recommend the apps to older adults	*"I would be willing to install it for my parents. They are more exposed to such scams (voice phishing) than we are, and might impulsively act out of concern for their family."* (P16)
	Willingness to install	*"Yes, I would. Because voice phishing affects emotions, and AI might not be influenced by emotions (compared to humans)"* (P16)
Age-based perceived vulnerability to threats	Children are vulnerable	*"So, children are exposed to mobile phones from a young age, right? Recently, I heard about a child who clicked on something while playing a game. It turned out to be something related to voice phishing, or so it seems. I didn't get the details, but it seems like it could be very dangerous for children, especially since they're getting phones at younger ages these days"* (P29)
	Older adults are vulnerable	*"The age group that most often falls victim to voice phishing is primarily the elderly."* (P6)
	Young generation won't fall for it	*"Most young people think they won't be victims. They think I won't be fooled because I'm quite smart... It's not that I have an aversion to it, but I genuinely believe I won't be a victim."* (P6)
Attitude towards on-device AI voice phishing detection apps	Effective countermeasure	*"I've taken a few calls briefly, and for a moment, I really thought, 'Huh? Could this be real?' It would be great if there were some sort of conclusive warning, like a cautionary text message or notification, to alert us."* (P14)
Experience with vishing	Acquaintance's experience	*"It wasn't me, but my mother who almost fell victim. She felt something was off and suspected it was voice phishing while on the call, but the fear of 'what if' prevented her from hanging up. I wish there was something that could help people calmly reassess the situation in moments like that."* (P21)
	Direct experience	*"I actually received a voice phishing call once and talked for a long time. I thought it was unbelievable how detailed they were with my personal information and even knew about my friend's situation in detail, which made me take the call. Despite having an anti-phishing app called whowho installed, it was ineffective in detecting it. If it could filter such (voice phishing) calls, that would be really helpful."* (P25)
Privacy concerns	Feeling like being wiretapped	*"I do have some worries. To be honest, if I exaggerate a bit, it might feel like being wiretapped."* (P17)
	Private information leakage	*"I know it only converts phone conversations to text on my phone, but I'm somewhat concerned about privacy, so I don't think I'll use it."* (P8)
Usage satisfaction	Feeling discomfort while calling	*"Before (the call), it was fine, but after receiving the call, it became extremely slow and laggy, which I found quite inconvenient."* (P21)
	Feeling discomfort, but not beyond usual	*"But honestly, the stuttering and heat were way less than I usually experience, so ..."* (P22)

F Usage satisfaction scores from the non-call survey, call-survey, and the difference between them.

Activity	**Usage satisfaction**	**Control group**			**Experimental group**		
		Non-call	**Call**	**Difference**	**Non-call**	**Call**	**Difference**
Web surfing & chatting	Overheating	4.67 ± 0.49	4.27 ± 0.96	-0.40 ± 0.74	4.67 ± 0.82	4.40 ± 0.74	-0.27 ± 0.59
	App freeze	4.47 ± 0.74	4.00 ± 1.20	-0.20 ± 1.21	4.87 ± 0.52	4.40 ± 1.12	-0.47 ± 1.19
	App crash	5.00 ± 0.00	5.00 ± 0.00	0.00 ± 0.00	5.00 ± 0.00	4.93 ± 0.26	-0.07 ± 0.26
	Overall sat.	4.33 ± 0.49	4.13 ± 1.19	-0.20 ± 1.26	4.53 ± 0.83	4.13 ± 1.19	-0.27 ± 0.88
Watching a video	Overheating	4.73 ± 0.80	4.53 ± 0.83	-0.20 ± 0.41	4.87 ± 0.35	4.53 ± 0.74	-0.33 ± 0.49
	App freeze	4.80 ± 0.56	4.53 ± 0.92	-0.27 ± 0.96	4.67 ± 0.49	4.60 ± 0.74	-0.07 ± 0.26
	App crash	5.00 ± 0.00	5.00 ± 0.00	0.00 ± 0.00	5.00 ± 0.00	5.00 ± 0.00	0.00 ± 0.00
	Overall sat.	4.73 ± 0.46	4.67 ± 0.62	-0.07 ± 0.46	4.73 ± 0.46	4.67 ± 0.62	0.00 ± 0.38
Playing a game	Overheating	4.40 ± 0.74	4.20 ± 0.86	-0.20 ± 0.56	4.73 ± 0.59	4.07 ± 0.88	-0.67 ± 0.62
	App freeze	4.67 ± 0.62	4.53 ± 0.83	-0.13 ± 0.64	4.93 ± 0.26	4.93 ± 0.26	0.00 ± 0.00
	App crash	4.80 ± 0.77	5.00 ± 0.00	0.20 ± 0.77	5.00 ± 0.00	4.93 ± 0.26	-0.07 ± 0.26
	Overall sat.	4.53 ± 0.83	4.47 ± 0.64	-0.07 ± 0.46	4.87 ± 0.35	4.87 ± -0.35	0.00 ± 0.38