A System overview of the prototype app

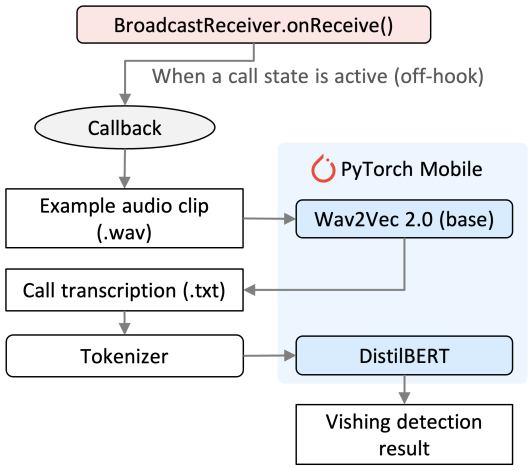

Figure 4: System overview of the prototype app.

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