

In-lab Interview: In the lab room

Hi, my name is *[name]*. I am *[self describe]*. Thank you for participating in our study. To ensure that we give everyone the same instructions and feedback, I will be using a script. Also, for the later data analysis, I am wearing a smart glass to record your interaction with the app we use. Is it OK to start recording now?

We know you have consented to the study. In this study, we aim to explore how blind individuals interact with and personalize visual question-answering AI technologies. For this purpose, we will be using the BeMyAI app. In this app you can simply take a picture and Be My AI will give you a detailed description of the picture. You can chat and ask further questions with Be My AI through the app to get more information. *<Do you have any questions so far?>*

Have you already installed the app on your phone? If not, it is totally okay, we can do that with you here.

Just a heads up, all these apps come with their own privacy policies that are beyond our control. We shared the link of the ByMyAI policy on your consent form, but we are also happy to read and discuss some aspects of their policy now before we start the study and app installation. Is it something you are interested in? Just a reminder that you can always stop participating in the study at any time. Just let us know.

If yes, one thing that is interesting about the policy is that

- You can ask [the app] to modify or delete your data by contacting [them] at legal@bemyeyes.com.

We recommend participants to do this at the end of the study as they wish.

[help the participant to install the app on their phone if they don't have it installed already.]

Including this session today, the study consists of 3 different stages. Today, session 1, we want to learn about you and also for you to try out some sample scenarios with the Be My AI app. This will help you become more familiar with the app and explore ways to personalize it. We estimate that this will take about 2 hours. *<Do you have any questions so far?>*

OK, let's start with some demographic questions.

<Demographics>

Now let's go over the app interface. Please open the Be My Eyes app and take a moment to explore. [Pause for participant to explore.] When ready, go to the "Be My AI" section at the bottom of the screen; this is where we'll be doing all interactions. [Pause for participant to open Be My AI.] Here, you can point your camera at what you want to capture and press the "Take Picture" button. The app will process the image, which may take a few seconds, and then give a description. You can then choose to take another picture or add a question or text.

Note that when you take the first picture in a session, you can't send a question with it. But after that, you can add text with each picture. Let me know if you have any questions about the interface.

Now, let's begin using the app with some sample scenarios. For each scenario, I'll provide context and an objective, along with suggestions on how to gather more information if needed. I will also ask some follow-up questions after each interaction.

Scenario 1 (getting familiar + sharing instructions):

Imagine your laptop is undergoing an OS update and you don't know if it has finished the update yet. You want to understand its update progress and whether it has reached the login page to enter your password, but your screen reader isn't working yet since you haven't logged into the main session. Use the Be My AI app to find out which screen your laptop is showing. Start by taking a photo of the screen, and the app will provide a description.

[Let the participant interact with the app]

For our research, we need to capture your interactions with the app. Let's go through the process step by step to show you how to share each interaction thread/session with us. *[Share interaction with us:*

- Add our email or phone number [301-405-2533] to their contact list.
- Practice sending the interaction to us (using email or message). [Use "copy" and then email]
- Prefer email or message for future communications?
- Let's practice again, this time on your own, with a second object.

Now, before we jump into some other scenarios, one thing we know about these apps is that they send the user photos to a server. So, it is not really private. Beyond avoiding using it with documents that include an id, banking, addresses, ssn, and medical stuff, you might also want to try avoiding including faces on the photos (either it is yours, a family member, a friend, or a passerby). This applies both for our scenarios here but also when you use the app on your own in the coming days.

Scenario 2 (Zero-shot: Style Prompting + binary feedback):

<Guide the participant to the hallway>

In this scenario, You're in an unfamiliar hallway on campus after finishing a meeting in someone's lab, like today. As you prepare to leave, you realize you turned around a bit and need to find that exit door, the one close to the elevator that you came in through. Of course there might be some students around, but you prefer to try to see how well the app works in helping you locate that door.

Now, let's just say if you didn't like how that really went and you want to let the app know so hopefully it gets better at it in the future. The simplest way to go about this though it might not be

the most effective one is to just give it a thumbs up or thumbs down. This app allows you to give a thumbs up or down at the end of the session, but it doesn't affect your current experience directly. you can type something like "not helpful", or "nope" or whatever. Similarly, if you really liked it you could type "great", "awesome" or whatever, and see how the app responds to that.

<Let the participant interact with the app to complete the task, one photo or multiple, follow up questions, etc, as they wish >

Q1: If you were the app and guiding yourself, what kind of responses would you give in this situation? Consider both the content (e.g., anything you'd add or remove) and the style (e.g., a friendlier tone, more subjective guidance, etc.).

<Let the participant give this type of feedback of their choice>

Another way you could help the app work better for you is to be more specific with instructions. For example, you can say give me more details/less details or only describe what is close to me. You can instruct it to be less verbose and give you the gist of it. These instructions can let you show the model what matters to you in this situation, helping to customize it to your needs, or in other words, personalize it. The model doesn't remember your preferences between sessions, but we're interested in learning how and when you'd use these personalized instructions options.

<Do you have any questions so far?>

<Let the participant interact with the app to complete the task>

[Share interaction with us]

Follow-up Question:

Q1: Did the app help you achieve your goal?

Q2: Can you tell us about these two techniques that we tried and how you would compare them?

Q3: How would you tell if the app's responses are right or wrong in this scenario?

[You can end the chat now]

Scenario 3 (Zero-shot: Role/Emotion Prompting):

In this scenario, imagine you are buying a snack from a local bottega. You are not familiar with some of the brands they carry. You pick one and want to check how many calories and protein it has to meet your goal. You try to use the app to tell you whether this is a good option.

<Let the participant interact with the app>

Q1: If you were the app and guiding yourself, what kind of responses would you give in this situation to better meet let's say a specific goal like low in calories while high in protein?

Consider both the content (e.g., anything you'd add or remove) and the style (e.g., a friendlier tone, more subjective guidance, etc.).

There's a lot of information on food packages, and sometimes you might just need some specific information. Besides the approaches above, another way you can improve the responses you get from the app is to mention what you're aiming for (express your intention of your question or this interaction). For example, one might say that they are trying to lose some fat, so they want to watch their calorie in-take, or they want to gain some muscle so they care about the amount of proteins in food. This way the model may adapt its response based on your input.

<Do you have any questions so far?>

<Let the participant interact with the app to complete the task>

[Share interaction with us]

Follow-up Question:

Q1: Did the app help you achieve your goal?

Q2: If you were to choose between the [last favorite] approach and this one with expressing your intention/goal, which one would you choose? Why?

Q3: How would you tell if the app's responses are right or wrong in this scenario?

[end chat]

Scenario 4 (Thought-Generation (Chain of Thoughts)):

Remember the [food] that we used in the first scenario? Now imagine you have left it somewhere in the lab room, but forgot where exactly. You can use the app to help you find it.

<Let the participant interact with the app for no more than two rounds>

Q1: If you were the app and guiding yourself, what kind of responses would you give in this situation to better help you locate the [food]? Consider both the content (e.g., anything you'd add or remove) and the style (e.g., a friendlier tone, more subjective guidance, etc.). How would you tailor the app's responses.

Again, you can always use some techniques to guide the app to better assist you. You can provide some examples or instructions to the app, and explain your reasoning behind them. For example, you can say: I only care about finding the [food]. So please do not tell me about anything that's not relevant to this process, such as the presence of a clock on the wall. In this example, asking the app not to mention about the clock on the wall is an example of what you do not want, and the part of "only caring about finding the food without any irrelevant information" is your reasoning. The combination of examples and explanations may help the app to better understand your needs and adjust its responses accordingly.

<Do you have any questions so far?>

<Let the participant interact with the app to complete the task>
[Share interaction with us]

Follow-up Question:

Q1: Did the app help you achieve your goal?

Q2: If you were to choose between the [last favorite] approach and this one with providing instructions or examples with reasoning, which one would you choose? Why?

Q3: How would you tell if the app's responses are right or wrong in this scenario?

Scenario 5 (few shot):

In this last scenario, imagine that you left your keys somewhere in the room but you can not remember where you left them. Just like before, you can use the app to help you find it.

Sometimes, AI models (like the app) might not identify objects perfectly. To improve its accuracy, you can give it a few examples or describe some features of your keys to help it recognize what you're looking for. For example, you could describe the appearance of your keychain and highlight specific features, like mentioning that the keys are attached to a green keychain with some text on it.

<Do you have any questions so far?>

<Let the participant interact with the app to complete the task>
[Share interaction with us]

Q1: If you were the app and guiding yourself, what kind of responses would you give in this situation to better help you locate your key? Consider both the content (e.g., anything you'd add or remove) and the style (e.g., a friendlier tone, more subjective guidance, etc.). How would you tailor the app's responses.

<Let the participant interact with the app to complete the task>
[Share interaction with us]

Follow-up Question:

Q1: Did the app help you achieve your goal?

Q2: If you were to choose between the [last favorite] approach and this one with providing examples of the object, which one would you choose? Why?

Q3: How would you tell if the app's responses are right or wrong in this scenario?

I hope this was a fun session. We tried out many different scenarios and some techniques to make this technology work better for a user. But this app doesn't really have a memory. So, it won't really improve from session to session. But the one we are interested in building, will. So basically, we are hoping that in the next 10 days that you will interact with the app, you can try some of these techniques where they make sense for you and share with us the interactions so that we can learn which approach can be more effective.

Diary Study:

3-days Practice:

Day 1: Set up to be successful.

Hey there, welcome to day 1 of our study! We're thrilled to have you on board. In the first three days, we have provided some instructions for you to get comfortable with the app, but feel free to extend beyond our instructions and play around with it. Then, for the following week, go ahead and use it in all sorts of daily situations that you think are fitted.

During the first day, try to ask the app to help you recognize items in your daily life environment, such as food and clothing. As you interact with the app, we encourage you to experiment with different techniques to improve the app's responses for you whenever you think it could help. You can try methods we discussed in the in-person session, such as indicating "helpful" or "not helpful," expressing your goal, providing examples with reasoning, describing the object you're looking for, or any other feedback you feel might improve the app's performance for you. Please send us at least three examples of your interactions with the app.

When sharing your interactions, please also send along some context, like your goal for the interaction and whether it worked, just like you did in the in-person session. Please also include details about the setting, for example, mention if it was outdoors, around other people, or in a familiar place. This will help us understand your experience better.

Day 2: Try lighting and disability things.

Great to see you back! For today's practice, let's play with lighting. For example, you can take photos in bright natural daylight, at night with all room lights on, or with just one light on in the darkness. Beyond that, you can also explore how well it works in recognizing accessibility objects such as mobility aid tools. As you interact with the app, we encourage you to experiment with different techniques to improve the app's responses for you whenever you think it could help. Please send us at least three examples of your interaction with the app. It'll be interesting to see how the app performs under these different conditions!

When sharing your interactions, please also send along some context, like your goal for the interaction and whether it worked, just like you did in the in-person session. Please also include details about the setting, for example, mention if it was outdoors, around other people, or in a familiar place. This will help us understand your experience better.

Day 3: Try some different environments. E.g. outdoor, or in a less-familiar environment.

Today is the last day of practice. During today's practice, we encourage you to step out of your comfort zone and try the app in different settings than you've used in the last two days. You can perhaps try some outdoor settings, some environments that you are less familiar with, or some settings where you are not alone. Whenever you think it may be helpful, feel free to try out different techniques to make the app work better for you. Please send us at least three examples of your interaction with the app. It's a great opportunity to explore new possibilities!

When sharing your interactions, please also send along some context, like your goal for the interaction and whether it worked, just like you did in the in-person session. Please also include details about the setting, for example, mention if it was outdoors, around other people, or in a familiar place. This will help us understand your experience better.

Day 4-10:

Over the next week, have fun exploring the app with all sorts of daily situations. We'd like to see at least 25 app interactions from you by the end of this 7-day period. You can spread them as you wish, for example, you can send four interactions every day or send nine interactions in any three days. Please send us your interaction sessions along with brief explanations of the interaction settings and log your daily experiences with the app by responding to our email/message.

As you interact with the app, we encourage you to experiment with different techniques to improve the app's responses for you whenever you think it could help. You can try methods we discussed in the in-person session, such as indicating "helpful" or "not helpful," expressing your goals, providing examples with reasoning, describing the object you're looking for, or any other feedback you feel might improve the app's performance for you. Please let us know if you have any questions.

Daily Reflection Questions:

Question 1: Was the app helpful today? Please explain.

Question 2: Did you try to give it any feedback or make it work better for you? Please explain.

Question 3: How could you tell if the app was right or wrong? Please explain.

Question 4: Anything else you would like to add?

Post interview:

Hello, my name is [name]. Thank you for participating in our final session. For the later data analysis, I will record this Zoom call. Is it OK to start recording now?

[Start recording when participants consent]

During this session, we will go over some of the data that you collected during the study and I will ask you some questions about your thoughts and experience.

So, we have chosen here a random sample of 6 interactions that you shared with us. In <3 (or less)> of those you attempted to tailor the app to work better for you. For the rest you didn't.

The first one was on <date> <time>. You capture a photo of what appears to be <a short visual interpretation> with the intent to <participant's shared intent>. The app (initially) said <app description>. You mentioned that you tried to make it work better for you by <participant's shared approach>.

Do you recall this case? (if not, provide a bit more context).

What would you have liked as an ideal response from the app?

What was your rationale for using <approach> instead of just <approach a> <approach b>?

Would you try something different now?

[repeat for each attempt to tailor up to three]

Now, for the rest of the examples, you didn't mention any effort to make the app work better for you. The first one was on <date> <time>. You capture a photo of what appears to be <a short visual interpretation> with the intent to <participant's shared intent>. The app (initially) said <app description>.

Do you recall this case? (if not, provide a bit more context).

What would you have liked as an ideal response from the app?

What was your rationale for not using any of the approaches to make it work better?

Would you try something different now?

That was a lot! Thanks for going down this memory lane with us. It helps us to better understand the experiences that you shared. We have xx more questions that are meant to capture your overall experience in this study and thoughts for the future.

For the first two questions, we will use a scale from 1 to 5, where 1 means "strongly disagree", 2 "disagree", 3 "neither disagree nor agree", 4 "agree", and 5 "strongly agree." Using this scale, please rate the following statements and share your reasoning:

- The app was overall easy to use. Please explain.
- I felt comfortable using the app. Please explain.

- Throughout the study you explored a few futuristic ways that you could make the app work better for you. Do you have any preferences among those approaches? Please explain. <prompt for follow-up questions like the ones they liked the least, and any challenges or the other way around>.

- In the future, what would be some of the cases where you feel like you could trust the app as is? <prompt for follow-up questions like the intent, context, etc>
- In the future, what would be some of the cases where you feel like you could put some effort into making the app work better for you?

- In the future, what would be some of the cases where you want the app's response to be more concise?
 - ... to contain more relevant information? Or to contain more detailed information?
 - ... to be more subjective or objective?
 - ... to be correct?
- What is the question that you would want us to have asked?

This is the end of the session. Thank you for participating. We really appreciate your time and feedback. We are paying [amount] to compensate for your time.

<Study Cleanup>

- Terminate the Zoom call.
- Retrieve all study materials.
- Copy and save the Zoom recording to a password secured server.
- Save the answers to questionnaires in a password secured server.