

Android

An Introduction to Seeing AI



This document has been designed to introduce you to the features found in the popular Smartphone app for Android, called Seeing AI, by Microsoft.

Assumptions Made in the Creation of this Document

A few assumptions have been made in the creation of this document. They are as follows:

- The latest version of Seeing AI has been downloaded and installed onto your Android Smartphone. At the time of writing (14 March 2025), the latest version is 1.3.
- You are familiar with navigating the various screens found on your Android Smartphone.
- Your Smartphone is connected to the internet.

Note 1: The instructions given in this document have been created using a Samsung Galaxy S21 Smartphone, running Android version 14. If you are using a non-Samsung device and/or a later version of Android, you may find some minor differences. However, it isn't anticipated that you will!

What is Seeing AI?

Seeing AI, by Microsoft, is a free app that narrates the world around you. Designed with, and for the blind and low vision community, this ongoing research project harnesses the power of AI (Artificial Intelligence) to open up the visual world.

What Can Seeing AI Do?

Seeing AI is a talking camera application, which can help with everyday tasks for people who are blind or have low vision. This takes the form of several features, known as channels, each with different intended functions. For example, the Short Text channel uses OCR (Optical Character Recognition) to read the text in-front of the camera, whereas the Scene channel uses computer vision to describe what's in a photo. The channels are discussed later in this document, but most of them use AI to help describe text, people and objects in the environment around the user.

Opening Seeing AI

To open Seeing AI, first locate its icon on the Home Screen and activate it. Upon doing this, the Seeing AI application will be opened.

An alternative method of opening Seeing AI, is to use Google Gemini (or the Google Assistant on earlier versions of Android). Simply invoke Gemini and say "open Seeing AI". Upon doing this, the application will be opened!

When Seeing AI is first opened, the Help page for the currently selected feature will be displayed towards the bottom of the screen. By default, this will be the Read feature. This page outlines how the feature works and also contains a link to a video of the feature in action. To close the help page, activate the "Close" button, situated at the upper right corner of the help page. Upon doing this, the main Seeing AI screen will be displayed.

Description of the Seeing AI Main Screen

At the upper left corner of the Seeing AI main screen, you will find the "Menu" button. When this button is activated, a number of options are available, including Help, Feedback and Settings. At the upper right corner of the main screen, you will find the "Quick Help" button, which when activated, provides an overview of the currently selected feature.

Just below the Menu and Quick Help buttons, you will find a row of three Tabs. These are labelled, from left to right, Read, Describe and More. When a Tab is selected, various controls are made available at the bottom of the screen. The Read Tab contains options for reading printed

and hand-written text. The Describe Tab allows you to take a photo and have it described. The More Tab contains a range of features, known as "Channels", including product recognition, currency recognition, colour recognition and light detection.

The area between the Tabs and the various controls, is dedicated to the text from documents, food items etc. This part of the screen takes up approximately 60% of the available screen space.

Following is a description of each Tab and its purpose. Most of the information has been taken directly from Seeing AI's help system!

The Read Tab

The Read Tab assists with printed text in the environment around you, as well as text from printed documents and handwritten notes.

When the Read Tab is activated, three buttons appear towards the bottom of the screen. At the centre of the screen, you will find the "Take Picture" button. To the left of the Take Picture button, situated at the left side of the screen, you will find the "Document Alignment" button. To the right of the Take Picture button, situated at the right side of the screen, you will find the "Recognition" button, which displays the current language being recognised.

There are two ways of reading text. Both of these are described as follows.

The Short Text Feature

This is the default feature. It allows you to hold the rear-facing camera on your device over some text and have it automatically read to you. As new text comes into view, it will also be read out loud. Whilst reading out loud, Seeing AI might start reading from the beginning again, if the camera captures a clearer view. This is known as recalibrating! For best results, hold your device approximately 12 inches away from the text. For larger text sizes, hold the device further away!

The main reasons for using the Short Text feature, are to find out quickly what a document is about, what a food item is or to find out what a sign on a door or wall says. It can also be used to quickly ascertain whether or not post is important or not.

The Short Text feature can be a little frustrating to use at times, due to the text potentially being repeated over and over. So it might be better to use what's known as the "Document" feature, described next, instead.

The Document Feature

To switch to the Document feature, activate the "Document Alignment" button, situated at the left side of the screen, to the left of the Take Picture button. Upon doing this, the Short Text feature will be temporarily disabled. You can now take a photo of the text for a more accurate description.

With the screen on your device facing you, the rear-facing camera will be at the upper right corner on the back of the device. Place it on the centre of the page containing the text, then slowly move it away. When Seeing AI has detected all of the text, it will say "hold steady" and a second or two later, a photo will be taken. Whilst the information is being processed, you will hear a series of tones. When the tones stop, the text from the document will be displayed on screen.

You can get Seeing AI to read the text aloud by activating the "Play" button, situated at the lower left corner of the screen. When the Play button is activated, the name of the button changes to become the "Stop" button. To stop the reading, activate the "Stop" button. Upon doing this, the speech will stop and the Stop button changes to become the "Play" button once more.

Note 2: When capturing an image of a document, Seeing AI will guide you with camera placement until all edges of the document are visible and the photo is taken. Make the necessary adjustments until you hear "hold steady".

In addition to having the text read from beginning to end, you can ask Seeing AI questions about the text in the document. To do this, activate the "Ask Seeing AI" button, situated at the lower left corner of the screen. For example, if you have taken a photo of a menu, you could ask the question "what are the desserts", by typing your question into the supplied edit box and then activating the "Send" button, which is situated at the bottom right corner of the keyboard. A few seconds later, Seeing AI will provide you with information specific to the Desserts only. Be aware that answers are AI generated, so mistakes are possible.

To read the AI generated information, activate the "Play" button, situated at the lower left corner of the screen. Remember, this button changes to be the "Stop" button when it is activated! To close the AI generated information, activate the "Back" button, situated at the upper left corner of the screen.

When you have finished with the current page, you can take a photo of another page by activating the "Add Page" button, situated near the upper right corner of the screen, just to the left of the "Quick Help button". Alternatively, you can close the document by activating the "Back" button, situated at the upper left corner of the screen.

Note 3: The Document channel works best when there is a high contrast between the page and the background, such as a white page on a dark surface. Sometimes, Seeing AI won't offer you suitable guidance when lining up the text. If this is the case, you can activate the Take Picture button to capture the text manually. If the page has no text on it, you will receive a message stating that no text was found.

Reading Handwriting

The Document feature also allows you to take a photo of handwritten text and have it read to you.

With the screen on your device facing you, place it on the centre of the page containing the handwriting, then slowly move it away. When Seeing AI has detected all of the text, it will say "hold steady" and a second or two later, a photo will be taken. Once the handwriting has been processed, the results will be displayed on screen in text format. You can get Seeing AI to read this text by activating the "Play" button, situated at the lower left corner of the resulting screen. To pause the speech, activate the Stop button, which is situated in the same place as the Play button.

The Handwriting feature is still in development and will undoubtedly improve in its accuracy as time goes by. For now, it's a useful feature, especially for reading things like Christmas and Birthday cards.

Note 4: the Handwriting feature requires the text to be the right way up! Recognition accuracy will depend on the handwriting style, which will vary greatly from person to person. Don't rely on this feature to read your Doctor's handwriting!

Changing the Voice and Reading Speed

One thing you might wish to do for the reading of text, is the change the voice, or adjust the reading speed. To change the voice and reading speed, follow these suggested steps:

1. Activate the "Menu" button, situated at the upper left corner of the screen. Upon doing this, a menu will be displayed.
2. Locate and activate the "Settings" button. Upon doing this, the Settings screen will be displayed.
3. In the Settings screen, locate and activate the "Text to Speech Engine" button. Upon doing this, a screen will be displayed from which you can choose your desired speech engine. Typically, these will be "Samsung Text-to-Speech engine" and "Speech Recognition and Synthesis from Google".
4. Activate the desired Speech Engine, then activate the "Back" button, situated at the upper left corner of the screen to save your choice and return to the Settings screen.
5. In the Settings screen, locate and activate the "Select Voice" button. Upon doing this, a list of available voices will be displayed.

The available voices will depend on which Speech Engine you chose. Typically, there will be a selection of male and female voices for a range of accents, such as English Australia, English United Kingdom, English United States and English India. You won't know what a voice sounds like until you use the Short Text or Document channel!

6. Activate the button for the voice you wish to use.
7. Activate the "Back" button, situated at the upper left corner of the screen, to save your choice and return to the Settings screen.

Now for the reading speed!

8. Locate the Speech Rate slider.

9. Slide the slider, until you have selected the desired speech rate. For reference, 30% is a pretty comfortable reading speed!
10. Finally, activate the "Back" button, situated at the upper left corner of the screen, to close the Settings screen and return to the main screen of the currently selected channel.

The Describe Tab

This tab features the latest Artificial Intelligence for describing the environment, including photos you take. You can also ask follow-up questions to get more details. This work is still experimental, so please use caution.

When the Describe Tab is activated, three buttons appear towards the bottom of the screen. At the centre of the screen, you will find the "Take Picture" button. To the left of the Take Picture button, situated at the left side of the screen, you will find the "Face Recognition" button. To the right of the Take Picture button, Situated at the right side of the screen, you will find the "Switch to Front Camera" button.

To take a photo, point the rear-facing camera on your device at the object or scene you want to be described. Then, activate the "Take Picture" button. Upon doing this, Seeing AI will process the information. Whilst the information is being processed, you will hear a series of tones. When the tones stop, Seeing AI will describe what you captured.

If you want to find out more about the contents of the image, activate the "Ask Seeing AI" button, situated at the lower left corner of the screen. Then, in the "Ask a question" text field, type your question and activate the "Send" button. A few seconds later, Seeing AI will provide you with more details about the image.

To close the Ask Seeing AI screen, activate the "Back" button, situated at the upper left corner of the screen. Upon doing this, you will be returned to the description of the picture you took.

Saving a Photo

You can save a photo, along with its description, by activating the "Save Photo" button, situated to the right of the Ask Seeing AI button. When you do this, you will be prompted to allow Seeing AI to use your location. Your location is used to find nearby routes to navigate, and is included when saving photos to your device. There will be a few locations to choose from! When you have selected a location, activate the "While using App" button. Upon doing this, you will receive a message stating that your photo was saved successfully.

Exploring a Photo

Seeing AI can break Photos down in to individual elements. This allows you to explore a photo, using one finger, to find out where certain things are in the scene.

To explore a photo, activate the "Explore" button, situated at the lower right corner of the screen. Upon doing this, Seeing AI will play a sound to indicate that the photo is being prepared for exploration. After the sound has stopped, you can explore the screen using one finger. As your finger passes over an element, Seeing AI will tell you what that element represents in the photo. As you navigate the screen, a high pitched "ding" sound will play when an element is encountered.

To close the Exploration screen, activate the "Back" button, situated at the upper left corner of the screen. Upon doing this, you will be returned to the description of the picture you took.

Sharing a Photo

It is possible to share a photo, or the description of that photo, with others. To do this, activate the "Share" button, which is situated to the left of the "Explore" button. Upon doing this, you will be prompted to choose what it is you want to share. The options are "Share Image" or "Share Text". Upon choosing the appropriate option, a list of names you have recently communicated with will be displayed, and beneath this list, a list of the most popular apps you can use to share the photo or description.

From the list, choose either Messages, Mail or WhatsApp (if you have it installed). Upon doing this, the photo or description will be attached to the outgoing communication. You can now compose your message and send it in the usual way. Upon doing this, you will be returned to the description of the picture you took.

To close the description of a photo, activate the "Back" button, situated at the upper left corner of the screen. Upon doing this, you will be returned to the main Seeing AI screen.

Note 5: Image descriptions and responses are AI-generated, so mistakes are possible.

The More Tab

When the More Tab is activated, a number of icons are displayed at the bottom of the screen. Each icon represents a channel. From left to right, the channels are: Product, Person, Currency, Colour and Light.

To choose a channel, simply activate the icon which represents the channel you want to use. If the channel you chose is one you haven't used before, the help page for that channel will be displayed. You can close this page by activating the "Done" button, situated at the upper right corner of the screen.

Following is a description of each channel available in the More Tab.

The Product Channel

The product channel is represented by an icon at the bottom of the screen, depicting a bar code. It allows you to recognise products based on various types of codes printed on the packaging, including bar codes. These are the most common type of product code, and may be found on the back or bottom of a container.

Some products, including some that are manufactured by Unilever, feature an accessibility enhanced QR code with a special border around it, which is easier to detect. These are found on the front of the packaging.

To scan a bar code, first activate the icon for the Product channel. Then, hold the rear-facing camera over the item and Seeing AI will guide you audibly with the camera placement until the code is detected.

Move your device over the product until you hear beeps that indicate a code is nearby. Starting further away and slowly moving the device closer works best. The faster the beeps, the closer you are to the code. In the case of accessibility enhanced QR codes, the distance will also be announced.

- When a barcode/QR code is detected, Seeing AI will say the product name.
- If additional information about the product is available, you can activate the More Info button, situated at the bottom centre of the screen.

To have Seeing AI read the additional information aloud, activate the "Play" button, situated at the lower left corner of the screen. To stop the reading, activate the "Stop" button, which replaces the Play button.

When you have finished with the Product channel, activate the "Back" button, situated at the upper left corner of the screen. Then, activate the "Close" button, also situated at the upper left corner of the screen.

The Person Channel

The Person channel is represented by an icon at the bottom of the screen, depicting a cartoonlike character. It allows you to scan your surroundings using the rear-facing camera on your device, to find out how many people are around you, how close they are and their facial expressions.

If you've taught Seeing AI to recognise a particular person, their name will be announced when they come into view. You can teach Seeing AI to recognise people by activating the Face Recognition button, situated at the left of the screen, just above the Read Tab. Upon doing this, the "Recognisable People" screen will be displayed. Next, locate and activate the "Add" button, situated at the upper right corner of the screen. You will need to take 3 photos of a person, at slightly different angles. When you've done this, you can assign a name to the photo! Finally, activate the "Back" button to return focus to the main screen.

Note 6: Always ask permission before taking a photo of a person's face. Additionally, ask if they mind if the photo is stored on your device and tell them the reason why you want to do this!

Seeing AI will also tell you audibly, the position of each face it detects, which can be helpful for centring a person in the camera's view. To take a photo, activate the "Take Photo" button, situated at the bottom centre of the screen.

If you want to take a Selfie, activate the "Switch to Front Camera" button, situated at the right side of the screen. Then point the front-facing camera at yourself, with your device around 12 inches away and activate the "Take Photo" button.

The Currency Channel

The Currency channel is represented by an icon at the bottom of the screen, depicting a US dollar bill. It allows you to recognise bank notes, in a wide range of currencies! The default currency should be British Pounds.

To find out the denomination of a bank note, hold the rear-facing camera over a bank note. When Seeing AI has recognised the value, it will announce it. If you want to recognise currency other than British pounds, activate the "Recognition" button, situated at the right edge of the screen. Upon doing this, a screen will be displayed consisting of a list of 19 different currencies. Choose the desired currency and start over again!

Currency recognition is pretty accurate, but sometimes it can make mistakes. As identifying bank notes accurately is very important, it is recommended that you have someone you trust confirm the value of each bank note if you are unsure!

Note 7: Seeing AI will not differentiate between real and counterfeit currency.

The Colour Channel

The Colour channel is represented by an icon at the bottom of the screen, depicting an artists colour pallet. It allows you to find out the perceived colour of an object. Be aware that this will depend on several

factors. Colours appear darker when there is less light, or if the object is in shadow; a white surface may appear slightly yellowish when the lights are on!

Point the rear-facing camera on your device at the object or article of clothing and Seeing AI will attempt to tell you the colour of the object/article of clothing. For best results, hold the device approximately 12 inches away!

The Light Channel

The Light channel is represented by an icon at the bottom of the screen, depicting a shining sun. It is designed to let you detect how much light there is in your environment, using tones emitted from your device. The pitch of the tone is based on how much your device sees. The more light there is, the higher the pitch of the tone!

Note 8: Be aware that using some of Seeing AI's features, such as the Short Text, Document and Colour Detection channels, requires use of the torch. This can greatly impact on battery life.

This is the end of the introduction to Seeing AI. If you want to find out more about Seeing AI, contact the author, Gary Eady, using the following email address.

technology@sightforsurrey.org.uk