

EVERYONE · 5 MINUTE READ

# Images, voice and files: giving AI more than text

AI assistants can read images, files and speech; they describe and summarise well, but misread fine detail, so check extracted facts and numbers.

## What it is

Modern AI assistants, such as Claude, ChatGPT, Microsoft Copilot and Gemini, can take in more than typed text. You can give them a screenshot, a chart, a photo of a whiteboard, a scanned form, a PDF or a spreadsheet. Many also accept **voice input**: you speak or upload a recorded voice note; some assistants turn speech into text, others listen to it directly.

Think of a quick, well-read colleague with slightly poor eyesight: they grasp the overall picture, but may misread a small figure.

## Why it matters

- It saves retyping: notes on a whiteboard or a printed form can become an editable draft in a minute.
- Long files can be summarised before a meeting.
- AI can misread small print, handwriting and chart values, and it may not say so.
- Screenshots and files often hold more than you meant to share, such as names, email addresses or figures in a corner of the screen.

## How to do it

1. **Check** that your organisation allows the tool to receive files, images or recordings, and what kind of data may go in.
2. **Crop or redact** images first: remove names, faces, account numbers and anything confidential that the task does not need.
3. **Choose** clear inputs: a sharp, well-lit photo, a text-based PDF rather than a blurry scan, a spreadsheet rather than a picture of one.
4. **Say** exactly what you want: "List every action item", "Extract the table into rows", "Summarise in five bullets".
5. **Ask** the assistant to mark anything it could not read clearly, rather than guess.
6. **Check** every extracted number, date and name against the original before you use it.
7. **Keep** the original file, so anyone can trace the result back to the source.

## Example

After a workshop, you photograph the whiteboard.

**Before:** you spend 20 minutes typing up the notes, or they stay on your phone.

**After:** you upload the photo and ask:

This is a photo of our workshop whiteboard on [topic].  
1. Turn it into tidy meeting notes with headings.  
2. List every action item with the owner's initials if shown.  
3. Mark any word you cannot read as [unclear] instead of guessing.  
Do not add anything that is not on the board.

You compare the notes with the photo, fix two [unclear] words and one misread initial, then share them.

## What AI does well with images and files:

- Describing a picture or diagram in plain words.
- Pulling text from a clear, printed document.
- Summarising long PDFs and reports.
- Explaining what a chart seems to show overall.

## What it does badly:

- Reading small, faint or handwritten text.
- Giving exact values from a chart (it estimates from the picture).
- Counting many items in a photo.
- Reading very long files or large spreadsheets in full: they may be cut short or partly read, so ask what it read.
- Identifying people. Never ask it to; photos of people are personal data and should not be shared unless your organisation's rules allow it.

## Common mistakes

- **Uploading a full-screen screenshot.** Fix: crop to the part you need; check for open emails, chat windows and customer details.
- **Copying chart values straight into a report.** Fix: get the numbers from the underlying data, not from the picture.
- **Trusting text from a scan or handwriting.** Fix: compare every name, date and figure with the original.
- **Sending a whole spreadsheet with customer data.** Fix: remove columns you do not need, or test with made-up data.
- **Forgetting that voice notes are recordings.** Fix: do not record other people without their agreement, and follow your organisation's rules on storing audio.

## Try it this week

1. Find a non-confidential printed page or a whiteboard photo with no names on it (3 minutes).
2. Upload it to an approved assistant and ask for the text plus a five-bullet summary, marking anything unclear (4 minutes).
3. Check the extracted text against the original and count the errors (5 minutes).
4. Try once more by voice: speak a short request instead of typing, and note whether it was quicker (3 minutes).

## Going further

- For a chart, ask for the trend, then confirm figures from the source data.
- With spreadsheets, some assistants calculate on the actual data; ask to see the method so you can recheck it.
- Ask for extracted tables as Markdown so you can paste them elsewhere (see guide 06).
- For long PDFs, ask for page numbers next to each point, then spot-check them.
- Voice conversations work well for thinking aloud; ask for a written summary at the end and check it (see guide 09).

## Key words

- **Voice input:** speaking to the assistant instead of typing; some assistants turn speech into text, others listen to it directly.
- **Extract:** pull specific information, such as a table or list, out of a file or image.
- **Redact:** hide or remove sensitive details before sharing.
- **Scanned document:** a picture of a page, which is harder for AI to read than a text-based file.
- **Personal data:** any information that can identify a person, including photos and voice.