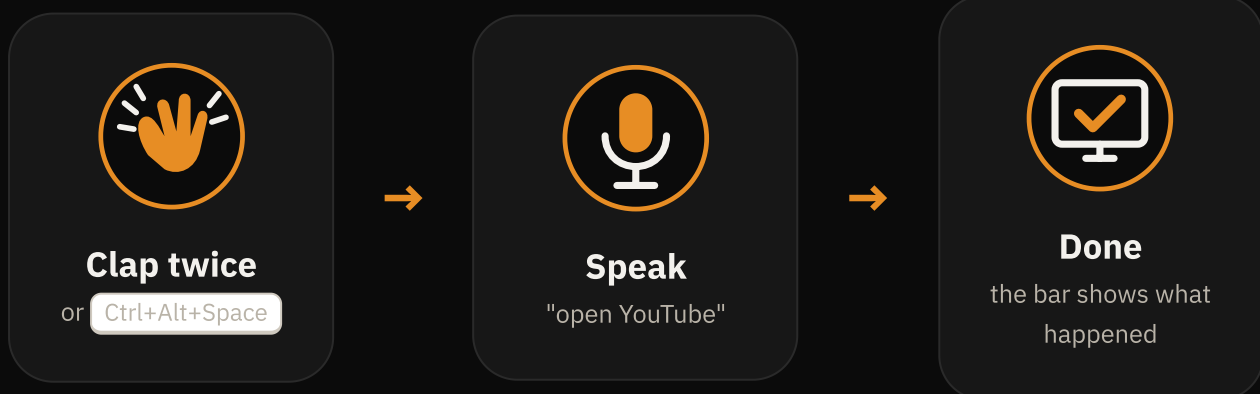


KaldrivoAI

User Guide

Talk to your computer in Arabic and English. Clap twice, say what you want, and your PC does it.

Version 1.6 · Windows 10 / 11



1 What is KaldrivoAI?

KaldrivoAI is a voice assistant for Windows that understands Arabic (including Gulf dialect) and English. Clap twice, a slim black-and-orange bar appears at the top of the screen, you say what you want, and the PC does it: opens apps and websites, plays YouTube, types for you, draws in Paint, turns web research into Excel tables, controls your Android phone and casts your screen to the TV. It runs on your PC and only ever performs a fixed list of safe actions.

2 Requirements

- **Windows 10 or 11.**
- **Python 3.10 to 3.13** from python.org. During setup, tick **"Add python.exe to PATH"**.
- **A microphone** (the laptop's built-in mic is fine).
- **Internet** for the first install and for speech recognition.
- **Optional:** a free Gemini key (section 5), a camera for the finger mouse, an Android 11+ phone for phone control, a TV that supports screen mirroring.

3 Install

- 1 Unzip `KaldrivoAI.zip` somewhere permanent (for example Documents). You get a folder called `KaldrivoAI`.
- 2 Open it and double-click `install.bat`. It sets up the program and installs its libraries (2–5 minutes the first time), then the finger-mouse libraries. If that last step fails, everything else still works.
- 3 Double-click `KaldrivoAI.bat` (or `run.bat`). The bar appears at the top of the screen. (If you skipped `install.bat`, it runs by itself.)
- 4 Allow the microphone: Settings → Privacy & security → Microphone → turn on "Let desktop apps access your microphone".

"Windows protected your PC"? That's expected: KaldrivoAI is a beta and isn't code-signed yet. Click **More info**, then **Run anyway**.

Future updates: quit (`Ctrl+Alt+Q`), unzip the new version over the same folder and choose **Skip** for `config.json` to keep your settings, then run `install.bat` and `run.bat`.

4 First run

Wake it

Clap twice, or press `Ctrl+Alt+Space`. The bar appears and listens.

Conversation mode

After the first command it keeps listening, so just say the next one. To close, say "close conversation" («سكر المحادثة») or press

`Ctrl+Alt+Space`.

Quit

`Ctrl+Alt+Q`, the tray icon → Quit, or say "quit assistant".

Move the bar

Say "move the bar" or use the tray's "Move bar", drag it, then double-click to lock it. "Reset bar" puts it back.

■ The tray icon (next to the clock)

Right-click the KaldrivoAI icon: Listen now · Phone QR · Move bar · Reset bar · Hand mouse · Phone camera... · Phone control... · Camera preview · Demo mode · Start with Windows · Voice replies · config.json · Log · Quit.

Want it to start with Windows? Tick "Start with Windows" in the tray menu; untick it to stop.

5 Gemini key (optional, free)

Each user adds their own key. The first time KaldrivoAI starts without one, the key window opens by itself; you can press "Later". All the basic commands work without a key. The key adds:

- Understanding of free, natural phrasing, and long sentences split into steps (from the same safe action list).
- Chat with a "Jarvis" persona and a natural voice; it remembers your name and things you ask it to keep.
- Research → Excel, screen description, and drawing anything you ask for.

- 1 Go to aistudio.google.com/apikey and sign in with a Google account.
- 2 Click **Create API key** and copy it (new keys start with `AQ.`, older ones with `AIza`).
- 3 In the key window (or run `set_gemini_key.bat`): Paste → pick your dialect from the list (optional) → Save.
- 4 If you opened the window from the file while KaldrivoAI is running, quit it with `Ctrl+Alt+Q` and start `run.bat` again.

Your dialect: with a key, KaldrivoAI replies in the Arabic dialect you speak. Saying "set my dialect to Egyptian" («لهجتي مصرية»); also Saudi, Gulf, Levantine, Iraqi... tunes speech recognition to your dialect and remembers it, so it understands you better. (New in 1.6)

6

Command cheat-sheet

Every example here was checked against the program. Chain commands in one sentence with "and / then", e.g. "open notepad and type hello world then take a screenshot". With the Gemini key you can also say things your own way.

Apps, web & quick info

Say in English	...or in Arabic	What happens
open Word	افتح الورد	Opens a program — Word, Excel, Paint, Notepad, Calculator, Settings or any installed app
open YouTube	افتح يوتيوب	Opens a website
open downloads	افتح التنزيلات	Opens a folder (Downloads, Documents, Desktop, Pictures...)
—	افتح ملف الميزانية	Finds a file by name and opens it
search for coffee shops in Riyadh	ابحث عن مطاعم في الرياض	Google search
open ChatGPT and ask it for three dinner ideas	افتح شات جي بي تي واسأله عن ثلاث أفكار للعشاء	Opens the site, types your question and sends it (ChatGPT, Gemini, Claude, Perplexity, Copilot, Grok, DeepSeek)
what time is it	كم الساعة	Tells the time
weather in Dubai	وش الطقس في جدة	Weather
what's on my screen	وش في الشاشة	Describes the screen (needs the Gemini key)
close the window	سكر النافذة	Closes the active window
close YouTube	سكر اليوتيوب	Closes only that site's tab, not the whole browser
minimize	صغر النافذة	Minimizes the window

YouTube & media

Say in English	...or in Arabic	What happens
search YouTube for lofi music	ابحث في يوتيوب عن سورة الكهف	YouTube search results
—	افتح اليوتيوب وابحث عن سورة البقرة وشغل أول فيديو	Plays the first video, hands-free
open the MrBeast channel on YouTube	افتح يوتيوب على قناة مستر بيست	Opens a channel
pause	وقف الفيديو	Pause / play
next video	الفيديو اللي بعده	Next

Volume, brightness & screen

Say in English	...or in Arabic	What happens
volume up	ارفع الصوت	Volume up
volume down	وطي الصوت	Volume down
set volume to 30	خل الصوت 50	Sets an exact level
mute	كتم الصوت	Mute
brightness up	زيد الإضاءة	Brighter screen
brightness down	قلل الإضاءة	Dimmer screen
set brightness to 70	خل الإضاءة 70	Exact brightness
take a screenshot	صور الشاشة	Screenshot

Typing, dictation & keys

Say in English	...or in Arabic	What happens
type hello world	اكتب السلام عليكم	Types the text where the cursor is
open Notepad and type hello	افتح المفكرة واكتب مرحبا	Opens the app and types into it
start dictation	ابدأ الإملاء	Everything you say is typed
stop dictation	وقف الإملاء	Stops dictation
press enter	اضغط إنتر	Presses Enter
copy / paste / select all	انسخ / الصق / حدد الكل	Copy, paste, select all
undo	تراجع	Undo
new tab / refresh the page	تبويب جديد / حدث الصفحة	Browser keys
switch window	بدل النافذة	Switches window
scroll down	انزل تحت	Scrolls
click on sign in	اضغط على تسجيل الدخول	Finds a button or link by its name and clicks it
right click	كلك يمين	Right-click

■ Drawing in Paint

Say in English	...or in Arabic	What happens
draw a camel	ارسم سيارة	Draws with the real mouse, then colours (20–35 s)
open Paint and draw a car	افتح الرسام وارسم مسجد	One of 12 built-in drawings
draw anything	ارسم لي أي شيء	A random drawing
draw a circle	ارسم دائرة	A simple shape

■ Files & saving

Say in English	...or in Arabic	What happens
save	احفظ	Saves the open file
close Paint without saving	سكر الرسام بدون حفظ	Closes without saving
close Notepad and save	سكر المفكرة واحفظ	Saves to Documents with an automatic name, then closes
create a Word document about the trip	اعمل ملف اكسل فيه أسماء وأرقام	New Word or Excel file
organize my downloads	رتب التنزيلات	Sorts files into folders by type
undo organize	تراجع عن الترتيب	Puts the files back

■ Reminders, calendar & email

Say in English	...or in Arabic	What happens
remind me in 10 minutes to call mom	ذكرني بعد ١٠ دقائق اتصل على أمي	Reminder after a delay
set a timer for 5 minutes	نبهني بعد ساعة	Timer
add an appointment tomorrow at 5 with the dentist	خط لي موعد بكره الساعة ٥ مع الدكتور	Calendar event (Google Calendar by default)
send an email to Sara about the trip saying we land at 6	ارسل ايميل لأحمد موضوعه الاجتماع وقل له بتأخر شوي	Prepares the email in Gmail; it is sent only when you say "send"

■ Research → Excel (needs the Gemini key)

Say in English	...or in Arabic	What happens
research the best laptops of 2026 and put it in an Excel file called laptops	ابحث في النت عن أفضل المطاعم في الرياض وحطها في ملف اكسل باسم مطاعم الرياض	Searches the web, builds a styled table with a Sources sheet, opens it (10–60 s)
research electric cars and put it in Excel	سولي جدول اكسل عن أحدث السيارات الكهربائية	Columns are chosen to fit the topic

■ The assistant itself

Say in English	...or in Arabic	What happens
—	غير اسمك إلى جارفيس	Renames the assistant
close conversation	سكر المحادثة	Ends the conversation; waits for two claps again
set my dialect to Egyptian	لهجتي مصرية	Tunes speech recognition to your dialect (Saudi, Gulf, Egyptian, Levantine...)
quit assistant	سكر المساعد	Quits the program
move the bar	حرك الشريط	Drag the bar elsewhere; double-click to lock it
remember that my meeting is every Sunday	تذكر إن اجتماعي كل أحد	Keeps a fact about you on your PC (needs the Gemini key)

■ Power (always asks "are you sure?")

Say in English	...or in Arabic	What happens
shutdown the computer	أطفي الجهاز	Shut down
restart the computer	اعد تشغيل الجهاز	Restart
put the computer to sleep	خل الجهاز ينام	Sleep
lock the screen	قفل الشاشة	Lock

■ TV, phone remote & camera

Say in English	...or in Arabic	What happens
cast to TV	اعرض الشاشة على التلفاز	Mirrors the laptop to the TV
stop casting	وقف العرض	Stops casting
connect phone	اربط الجوال	Shows the QR code for the phone remote page
start hand mouse	شغل الماوس بالإصبع	Finger mouse via the camera
stop hand mouse	وقف الماوس بالإصبع	Turns it off
hide preview	اخفي المعاينة	Hides the small camera window
use the phone camera	استخدم كاميرة الجوال	Phone camera instead of the laptop's

Phone control (v1.6)

Say in English	...or in Arabic	What happens
show my phone	افتح جوالي	Your phone's screen in a window you control with the mouse
close my phone	سكر جوالي	Closes the window
call dad	اتصل على أبوي	Calls straight from your phone (uses your phone contacts)
call 0501234567	اتصل على 0501234567	Calls a number
hang up	سكر المكالمة	Ends the call
open WhatsApp on my phone	افتح واتساب بالجوال	Opens an app on the phone
go back on my phone	ارجع بالجوال	Back button
go home on my phone	الشاشة الرئيسية بالجوال	Home screen
volume up on my phone	ارفع صوت الجوال	Phone volume (and "volume down on my phone")
take a screenshot of my phone	صور شاشة الجوال	Saved to Pictures on the laptop
save number dad 0501234567	احفظ رقم أبوي 0501234567	Saves a name and number in the assistant
sync contacts	حدث جهات الاتصال	Re-reads the phone's contacts
type hello on my phone	اكتب بالجوال مرحبا	Types on the phone (Arabic only while the window is open)
set up phone control	ربط التحكم بالجوال	Opens the phone-control setup window

7 Your phone as a remote

- 1 Put the phone and the laptop on the same Wi-Fi.
- 2 Say "connect phone" (or tray → Phone QR). A QR code appears.
- 3 Scan it with the phone's camera and open the link. (If Windows asks about network access, choose Allow.)

Control tab

- Big mic button: speak from the phone.
- A text box that sends any command.
- Quick buttons: volume, brightness, play/pause, next, screenshot, YouTube, desktop, lock, TV, 📱 my phone.

Screen tab

- The laptop screen, live.
- Tap = click, double tap = double-click, long press = right-click, long press and move = drag.
- Rotate button for full screen; a key row: scroll, Enter, back, switch window, backspace, Esc, copy, paste.

Keyboard mode

Turn on «**کیبورد**» (**keyboard**) (it turns orange): whatever you type on the phone goes into the field that has focus on the laptop. Extra keys appear: Enter, delete letter, delete word, copy, paste, Tab. An empty send = Enter; up to 2,000 characters per send.

Click the field on the laptop first. Password fields often block pasting.

8 Finger mouse

Say "start hand mouse" or use the tray. It uses the laptop camera and shows a small preview window.

Raise index finger, move	The pointer moves
Pinch thumb + index	Click
Pinch, hold 1 s, move	Drag; open your fingers to drop
Thumb + middle finger	Right-click
Two fingers up, move	Scroll
Open palm for 1 s	Pause / resume

● Use your phone as the camera

- 1 Install **IP Webcam** on your Android phone, open it → **Start server** (same Wi-Fi).
- 2 Run `set_phone_camera.bat` (or tray → Phone camera..., or say "use the phone camera").
- 3 Enter the address shown at the bottom of the app (it starts with `http://` and ends with `:8080`) and save.

9 Phone control setup (v1.6)

A one-time setup. You need an Android 11+ phone on the same Wi-Fi as the laptop.

- 1 **Enable Developer options:** Settings → About phone → tap "Build number" 7 times.
- 2 **Turn on Wireless debugging:** Settings → System → Developer options → Wireless debugging. (Names vary a little between phones.)
- 3 **On the laptop** run `setup_phone_control.bat` (or tray → Phone control..., or say "set up phone control"). The first time it downloads the phone tools (about 11 MB).
- 4 **On the phone** open Wireless debugging → "Pair device with pairing code". Type the IP:port and the 6-digit code into the «ربط الجوال» window and press «إقران» (Pair).
- 5 Type the IP:port shown on the main Wireless debugging screen and press «اتصال» (Connect). A  means it's saved.
- 6 Try it: "show my phone" («افتح جوالي»).

Calls: "call Ahmed" looks the name up in your phone's contacts; if several match it asks «أي واحد؟» (which one?). Calls go out right away, without a confirmation. Numbers you save by voice ("save number dad 05...") take priority.

Android 10 or older, or the phone is the Wi-Fi hotspot? Turn on USB debugging, connect the cable once and choose «عن طريق كابل USB» (via USB cable) in the setup window.

Typing Arabic on the phone works only while the «جوالي» phone window is open; English works any time.

10 Cast to the TV

- 1 Turn the TV on and make sure it supports screen mirroring (Miracast).
- 2 Say "cast to TV" («اعرض الشاشة على التلفاز») or tap TV on the phone page. KaldrivoAI opens Windows' `Win+K` panel.
- 3 To stop, say "stop casting".

The first time, the `Win+K` list opens and you pick your TV. To have it picked automatically, put your TV's name as it appears in the `Win+K` list in `config.json` under `"tv_name"`.

11 Troubleshooting

It doesn't hear me	Check the microphone permission (section 3). Run <code>list_mics.bat</code> and put the mic's number or part of its name in <code>config.json</code> under <code>"mic_device"</code> .
Claps not detected	Use <code>Ctrl+Alt+Space</code> meanwhile. In <code>config.json</code> → <code>"clap"</code> : not detected → lower <code>threshold</code> to 0.12–0.18; wakes too easily → raise it to 0.3–0.45.
Something failed	Run <code>run_debug.bat</code> : errors show in the window, plus the last lines of <code>voice_assistant.log</code> . Or tray → Log.
"Gemini didn't answer"	Check the internet and try again. Wrong key → set it again with <code>set_gemini_key.bat</code> . Error 404 = that model was retired; the program tries other models by itself. Error 429 = free quota used up; wait and try later.
Excel says «بدون بحث مباشر»	Your key wasn't allowed live search, so the table comes from Gemini's own knowledge. Check the numbers before relying on them.
Phone disconnected	Wireless debugging switches itself off after a phone restart or a Wi-Fi change. Turn it on again. KaldrivoAI retries and looks for the phone on the network; if that fails, run phone setup again.
Phone page won't open	Both devices on the same Wi-Fi; allow the program if Windows Firewall asks.
Paint: no colours / thin lines	Maximize Paint and run <code>calibrate_paint.bat</code> once.
"Python was not found"	Reinstall Python and tick "Add python.exe to PATH".

12 Privacy & safety

- **Whitelisted actions only:** KaldrivoAI performs a fixed list of safe actions and never runs system commands from what you say. Plans from Gemini are checked against the same list.
- **It asks first:** shutdown, restart, sleep and lock ask "are you sure?"; an email is only sent when you say "send". Phone calls go out straight away and the log keeps only the last 4 digits.
- **What goes to Google:** your voice after you wake it (for speech-to-text). With a key: the text of commands that need understanding, chat, research topics, and a screenshot only when you ask "what's on my screen". Spoken replies are voiced by Gemini or Microsoft's online voices.
- **What stays on your PC:** memory (`assistant_memory.json`), your key and settings (`config.json`), the finger-mouse camera image. The phone page and live screen travel only over your own Wi-Fi.

Questions or feedback? anasdesign1990@gmail.com