

Speech to Text User Guide

Table of Contents

User Guide.....	1
Evaluation Licences.....	1
Permanent Licences.....	2
Using the Desktop App.....	2
Using a Web Browser.....	2
Speech to Text Models.....	3
Dictation App.....	3
Game Chat App.....	4
Transcribe File App.....	5
Technical Information.....	5
Hotkey Setup.....	5
GPU Preferences.....	6
Transcribe File GUI.....	6
Optimal Privacy Setup.....	6
Administrator Override.....	7
Known Issues.....	7
Troubleshooting.....	8
Uninstalling.....	8
Disclaimers.....	8

User Guide

Since you are reading this document, you have probably already installed the software.

Launch one of the Start menu shortcuts: “Live Dictation”, “Game Chat”, or “Transcribe File”.

Evaluation Licences

If your computer has an internet connection, click **Yes** when prompted with “Allow online activation?”. On the next screen, click “Evaluate”.

If the computer is offline, click **No** when asked “Allow online activation?”. Copy the computer ID by clicking the Copy button and paste it into a text file. Transfer this file to another computer with internet access using a USB flash drive or another suitable method. On the internet-connected computer, paste the Computer ID into the form on the following web page <http://vrmac-softver.com/speech2text/evaluation> and click “Submit”. The server will generate a ZIP file containing the licence for your computer. Save this file to the USB flash drive and transfer it back to the offline computer. On the offline software activation screen, click “Import...” and open the ZIP file from the flash drive.

Permanent Licences

If you wish to continue using the software after the 30-day evaluation period has expired, you will need to purchase a licence. A registered user account with a verified e-mail address is required before you can do so. Most steps can be completed using either a web browser or the desktop app.

Using the Desktop App

Launch LicenceHelper.exe. On the Online Software Activation screen, click Register, enter e-mail and account password (display name is optional), click Submit.

Verify your e-mail address by clicking the link sent from noreply@vrmac-softver.com

Return to the Online Software Activation screen in LicenceHelper.exe, click Sign In, enter e-mail and account password, click Submit.

Click Purchase. Your web browser will open at <https://vrmac-softver.com/speech2text/order> Select your licence type and click Submit, then review and complete your payment on the next page. Shortly after the transaction completes, you should receive an automated e-mail.

Activate your licence. Switch back to LicenceHelper.exe, press the refresh button. An Activate button will now appear, click it to complete activation.

Using a Web Browser

Create a user account at <https://vrmac-softver.com/signup>

Verify your e-mail address by clicking the link sent from noreply@vrmac-softver.com

Sign in at <https://vrmac-softver.com/signin> Purchase a licence at <https://vrmac-softver.com/speech2text/order> Select your licence type, click Submit, then review and complete your payment on the next page. Shortly after the transaction completes, you should receive an automated e-mail.

Activate your licence — this step requires physical access to the computer on which the software will be used.

- Launch LicenceHelper.exe and open the Offline Software Activation screen.

- Click Copy to copy your computer ID, paste it into a text file, and transfer the file to a computer with internet access (for example, using a USB flash drive).
- On the internet-connected computer, navigate to <https://vrmac-softer.com/user> and sign in.
- In the Purchase History table, click the Activate link.
- Paste your computer ID into the form and click Submit. The server will generate a ZIP file containing your licence.
- Save the ZIP file to your USB flash drive and transfer it back to the offline computer.
- On the Offline Software Activation screen, click Import... and open the ZIP file to complete activation.

Speech to Text Models

Newly installed software is useless without speech to text model files. I distribute 15 of them for completeness, but I only recommend 3.

- Voxtral Mini BCML — for graphics cards with 4 GB VRAM.
- Voxtral Mini BCML text — for graphics cards with at least 6 GB VRAM.
- Whisper Turbo BCML — for game chat frontend: uses much less resources than Voxtral.

If you have purchased the software but unable to download models due to some networking shenanigans¹, please contact support. I have alternative methods to distribute model files but these are not automated, require some manual intervention.

If you want to copy-paste models across computers to save time and network bandwidth, they are stored in `%LOCALAPPDATA%\Speech to Text\Models`, feel free to copy-paste these *.cgm1 files. Paste the model files into that location on the destination computer, then click “Download models...”. The downloader will verify the file hashes to ensure integrity, and the newly added models will become available in all speech to text apps I ship.

The downloader tool does not verify software licences. If you wish to download models for use on another computer, install the software and run ModelDownloader.exe from the installation folder. Download the required models, then copy them to the computer where you want to use speech-to-text. Note that many USB flash drives use the [FAT32](#) file system which cannot store files exceeding 4 GB. Some of the models I ship exceed that limit.

The downloader tool stores partially downloaded content on disk and automatically resumes incomplete downloads on launch.

Dictation App

The application continuously transcribes live speech captured from the microphone and sends the final output to third-party text editors.

¹ Model downloader utility is based on [BitTorrent](#) protocol.

After loading a model, use the next screen to select the audio input device, specify the language, and test speech to text by clicking “Capture”.

On the following screen, assign a global hot key, optionally restrict allowed text destinations with checkboxes, and click Next. The application will then switch to a small, always-on-top status window. When idle, this window can be moved by clicking and dragging it with the mouse.

Alt+Tab to the text editor where you want the text, press the configured hotkey to start the dictation session, and begin speaking. That status window shows newly transcribed text as it becomes available. Press the same hotkey again to stop the session.

If you alt+tab away from the text editor to another window, or if the input focus is stolen by a rogue pop-up, the application will stop sending text to the editor but will continue transcribing your speech. In this situation,

the newly transcribed text is displayed in green colour, and the transcribed text is held in a buffer, as shown in the screenshot to the right. When this occurs, stop the



dictation session using the same hotkey — the status window will become interactive again, allowing you to copy the buffered text directly from it. Alternatively, pressing the transcription hotkey once more will flush the entire buffered text into whichever text editor currently holds the keyboard focus.

As with keyboard input, dictated text replaces the current selection. If you select all content in the editor with Ctrl+A, press the dictation hotkey, and begin speaking, the existing content will be replaced.

While a dictation session is in progress, the status window does not respond to mouse input. To move the window or exit the application, stop the session first.

Game Chat App

The application transcribes individual spoken sentences and places the resulting text on the Windows clipboard.

The configuration steps are similar to those for dictation: select and load a model, test speech to text, and set up a global hotkey. Clicking Next transforms the tool into a small circular window with a microphone icon, as shown in the screenshot to the right.



By default, you can move that window around the screen using the mouse. If you right click the circle and select “Disable mouse events” from the context menu, the window will no longer handle mouse input and clicks will pass through to whatever is behind it. To re-enable mouse input or exit the application, right click the system tray icon that appears after selecting “Disable mouse events”.

Once you are satisfied with the position of the status indicator window, press the global hotkey and start speaking. Press the same hotkey again to stop recording and begin transcription. Alternatively, you can simply stop speaking. At the time of writing, the default setting for maximum silence duration is 1.5 seconds, this means that as soon as the app detects a continuous 1.5 seconds of audio without speech, transcription starts automatically. After a short delay, a pop-up displaying the transcribed text will appear. By the time this pop-up is shown, the text has already been copied to the Windows clipboard and can be pasted into other applications.

Unlike dictation, the game chat app is designed to run alongside other GPU demanding software. Once setup is complete, the user interface is minimal, and the speech-to-text engine uses significantly fewer GPU resources than dictation.

Transcribe File App

The application transcribes audio files from disk with the same speech to text engine. The workflow is quite similar: load a model on the initial screen, test on the next screen, transcribe files on the final one.

Technical Information

Hotkey Setup

Both Dictation and Game chat applications trigger speech to text with a hotkey. And there are two mutually exclusive ways to specify that hotkey: keyboard hotkey, and raw input.

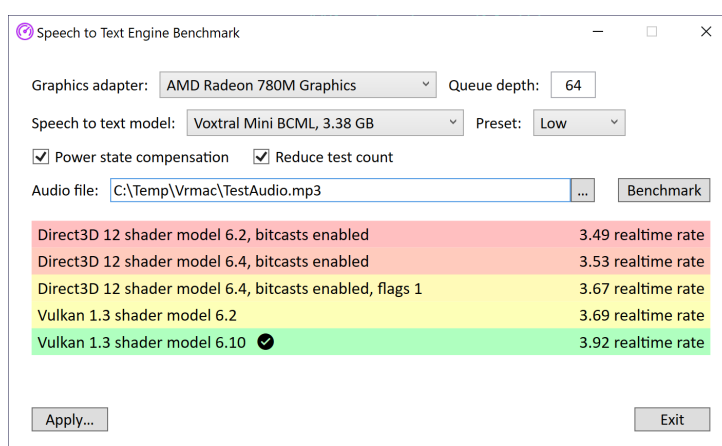
- Keyboard hotkeys rely on the corresponding [OS API function](#). The registered hotkeys are not specific to a particular keyboard device, and the hotkey device is never lost even if you disconnect all keyboards from the computer. If the hotkey you want is already registered by some other application running on the same PC, you will see an error message and the application will require you to specify another hotkey.
- Raw input supports mouse in addition to keyboard, and commands from other [human interface devices](#) like game controllers. The registered command is specific to the input device. Reconnecting the input device into another USB port changes device ID, you will have to set up the hotkey again. If the hotkey is already handled by some other running program, you will see no warning messages, the OS will deliver input events to both programs. And if you will use a keyboard-based raw hotkey, some security software might falsely flag dictation and game chat programs as keyloggers².

² If you are wondering why I support raw keyboard hotkeys despite the drawbacks, it is because some high-end mice present themselves to the operating system as composite USB devices, with one endpoint being a fake keyboard. When pressed, certain additional buttons beyond the five standard ones (left, right, middle, X1, X2) generate keyboard events.

GPU Preferences

All 3 applications have “GPU Settings...” button on the initial screen where you select a model to use. If you have a computer with two GPUs integrated and discrete, the integrated one is performant³, and you want discrete GPU for some other demanding tasks running in parallel with speech to text, switch the “Graphics adapter” preference.

I do not recommend using window for advanced preferences, though. Instead, run Benchmark.exe located in the installation folder. That tool transcribes a test audio file with the supported combinations of advanced parameters, and has an “Apply...” button to save the found optimal settings for use by speech-to-text applications.



Transcribe File GUI

When using Whisper models, the “Advanced options...” button on the test screen opens a window where you can adjust the speech-to-text engine parameters. These parameters are stored in the registry and are applied not only to the test screen, but also to production use. The default settings should be suitable for most use cases. Adjusting them may either reduce or introduce hallucinations in the output text, especially for low quality audio or non-speech parts of the audio.

Voxtral models are more reliable in my experience, hence the lack of the advanced options.

Optimal Privacy Setup

If you are transcribing very privacy sensitive audio (for example, you are a lawyer or a university professor), here’s the optimal setup.

In the GPU backend preferences screen, expand advanced options. Set “Diagnostic log files” setting to Disabled, which is the default. Find the “Process crash dumps” setting and **uncheck** “Save Minidumps” checkbox, which is checked by default. Press OK, close and reopen the software.

³ On my desktop PC, the integrated Radeon 780M GPU delivers 3.9 realtime rate with Voxtral BCML model.

Administrator Override

There's another way to set these two settings for all users in the system.

Launch Registry Editor. Windows will prompt for administrator permission; accept the prompt. Navigate to `HKEY_LOCAL_MACHINE\SOFTWARE\Vrmac Softver\Speech2Text` creating the key if it doesn't exist.

Right-click on the right-hand panel and choose New / DWORD (32-bit) Value. Name it `disableCrashDumps` with no spaces or quotes, and set the value to 1. Decimal or hexadecimal doesn't matter, they agree for very small integers. Create another DWORD value `disableFileLogger` with no spaces or quotes, set the value to 1.

The registry settings from `HKEY_LOCAL_MACHINE` disable the corresponding settings in the GPU backend preferences screen, and they require an administrator permission to modify.

These settings are applied when the software launches. If any of the transcribe apps were running while you did that, please close and reopen them for the change to take effect.

Known Issues

I haven't tested the software with Intel GPUs. All three GPU backends (D3D11, D3D12, and Vulkan) are vendor agnostic in principle, and modern Intel GPUs, both discrete Arc and integrated ([Lunar Lake](#) and a few other current generations) should comfortably exceed the performance requirements based on their TFLOPS specs. Alas, I don't have any of these devices here.

If you are trying to transcribe OGG [Vorbis](#) audio file but it fails complaining about missing codec, please install [Web Media Extensions](#) from windows store.

If you are monitoring the GPU tab in Windows Task Manager while transcribing audio, you may notice that the GPU appears idle, even though the computer's fans are spinning and the system is clearly under load. This is a display issue in Task Manager. To correct it, select a different graph from the drop-down list. The appropriate graph depends on the GPU model and driver version. When using Direct3D 12 or Vulkan backends, GPU activity is shown on one of the graphs labelled "Compute <something>", such as "Compute 0" or "Compute_1". With the legacy Direct3D 11 backend, the transcription workload will instead appear on the default "3D" graph.

When using the dictation app to input text into OpenOffice Writer 4.1.1 or LibreOffice Writer 26.8.0.3, it may take two or three attempts to start a dictation session. I think this is due to an accessibility-related bug in both applications. At least on my computer, OpenOffice 4.1.16 works correctly on the first try.

Troubleshooting

The software should run on a wide range of GPUs, including older or virtualised ones. I've tested it and confirmed good reliability on nVidia [Lovelace](#), AMD [RDNA 3](#), and AMD [GCN 5.1](#). Other hardware, including untested architectures, may perform as well or better, but I can't vouch for it.

If it doesn't work on your system and you'd like to find out why, there are two controls in the advanced GPU preferences screen that may help. The diagnostic log files combobox enables text logs. Logging works on all systems, regardless of GPU; the logs are saved to %LOCALAPPDATA%/Speech to Text/Logs. The GPU crash dumps checkbox enables GPU-side failure reports, but it's only available when running on an nVidia GPU⁴.

Uninstalling

The installer supports two distinct installation types: system wide, and current user. You can even install both ways if you want, the setup program does not stop you.

To uninstall the system wide installation, open either control panel "Uninstall a program", or settings "Apps & features". Type or copy-paste "Speech to text" into the search box, click the uninstall button.

To remove the program installed for current user only, copy-paste the following into the address bar of Windows explorer: %LOCALAPPDATA%/Speech to Text Press Enter, and delete Binaries folder there. Then copy-paste the following into the address bar: %APPDATA%\Microsoft\Windows\Start Menu\Programs Speech to Text folder.

Disclaimers

The speech-to-text engine can make mistakes, including substituting words that are phonetically similar but have entirely different meanings. Always review transcribed text carefully before sending or publishing, especially for important communications.

See the "Pre-existing IP.txt" text file in the installation folder for copyright notices.

⁴ To implement that feature, the software uses [Nsight Aftermath](#). I have downloaded the SDK for free; my company is neither sponsored nor endorsed by NVIDIA.