How to Install the MedAi Console

System Requirements

Operating System: MacOS
Computer: Apple Silicon Mac (M Series)
RAM: 16GB
Free disk Space: 10 GB

Step-by-Step Setup

If you already have Ollama, UV and ffmeg installed then please skip those steps.

1. Download and install the Ollama desktop application

This is the link to download Ollama. After downloading, please install it on your computer.
Then launch it. A white chat window will open.
https://ollama.com/

Ollama launches automatically when you start your computer.

2. Download the MedGemma-1 GGUF model from Huggingface (7.77 GB)

---

1. Open the terminal on your Mac
2. Paste in this line and press Enter:

```
ollama run hf.co/unsloth/medgemma-4b-it-GGUF:BF16
```

Optional:
If you also want to download MedGemma-1.5 GGUF (7.77 GB), use this command:

```
ollama run hf.co/unsloth/medgemma-1.5-4b-it-GGUF:BF16
```

3. Install ffmpeg

Use Hombrew (https://brew.sh/).

1. Open the terminal on your Mac
2. Paste in this line and press Enter:
```
brew install ffmpeg
```

4. Install UV

Paste this command into the terminal and press Enter:
```
wget -qO- https://astral.sh/uv/install.sh | sh
```

5. Download the project folder and place it on your desktop

5.1 If downloading from **GitHub**:

   a. On GitHub click on "<> Code". Then select "Download Zip"
   b. Download the project folder and unzip it.
   c. Inside the main folder you will find a folder named:
      MedAi-Console-v1.0
   d. Place MedAi-Console-v1.0 on your desktop.

5.2. If downloading from the **Kaggle writeup**:

   a. Download the MedAi-Console-v1.0 project folder.
   b. Unzip it.
   c. Place the MedAi-Console-v1.0 on your desktop.

5. Install the Console
---

1. cd into MedAi-Console-v1.0 folder:
```
cd Desktop
cd MedAi-Console-v1.0
```

7. Paste this command into the terminal and press Enter:
(This overwrites the file and changes the file permissions to make it executable.)
```
cat start-mac-app.command > temp && mv temp start-mac-app.command && chmod +x start-mac-app.command
```

8. Open the MedAi-Console-v1.0 folder

9. Double click this file: start-mac-app.command

10. The app will auto download all requirements and then open in your browser.

Using the Console

The name of the model you downloaded will appear in the dropdown menu in the top left.
You can submit images and pdf documents in addition to text.

You can also talk to the AI model by clicking on the Mic icon.
The AI voice is turned off by default. You can turn it on in Voice Settings.
Any changes you make to the settings will be automatically saved.

The app does not stop running when you close the browser tab.
To shut down the app, close the terminal window.
You can also close the terminal by selecting it and typing Ctrl+C on Mac.

Future startup

Now that the setup is complete, in future simply double-click the start-mac-app.command file to launch the app.
The project folder must be placed on your desktop before the app is launched.

You could start the app and leave it running in the background all day.
Then whenever you want to use it, enter the following url in your browser:

http://127.0.0.1:5000/

Your browser will remember this local address so you won't have to.

Troubleshooting

- It will take a few seconds to get a response to your first message because Ollama needs to load the model. Subsequent responses will be faster.
- If the app doesn't start, make sure Ollama is running (look for its icon in your menu bar)
- If you see "connection refused", restart Ollama
- Inference time will increase as you increase the context size setting.
- Inference time will be longer when submitting large images or pdf files.
- For the voice (TTS) to work Kokoro needs two files: kokoro-v1.0.onnx, and voices-v1.0.bin. These files are auto downloaded during the setup process. However, if the voice is not working then please download these files manually and place them in the project folder:

  kokoro-v1.0.onnx:
  https://github.com/thewh1teagle/kokoro-onnx/releases/download/model-files-v1.0/kokoro-v1.0.onnx

  voices-v1.0.bin:
  https://github.com/thewh1teagle/kokoro-onnx/releases/download/model-files-v1.0/voices-v1.0.bin

---
Notes

- When creating agents/tools that will generate math notation, you need to tell the agent to use LaTeX when generating math notation. Please add this note to the system message: Use LaTeX notation for mathematical or scientific expressions only.

- For best results when using your voice, use a headset or earphones with a mic. This reduces background noise. It also allows for a more relaxed chat because you won't have to constantly focus on being clearly heard by the speech to text system.

- Because MedGemma-1.5 is a reasoning model, I've set up the app to disable voice output when the MedGemma-1.5 model is selected.