Hier ist ein lauffähiges Minimalbeispiel (PySide6-GUI mit Textfeld und „Mikrofon“-Button). Aufnahme → Transkription mit **faster-whisper** (Whisper large-v3). Läuft lokal, ohne Cloud.

**Downloads:**

* [main.py](sandbox://mnt/data/stt_ollama_gui_demo/main.py)
* [requirements.txt](sandbox://mnt/data/stt_ollama_gui_demo/requirements.txt)

**Installation (Windows, nicht WSL)**

Mikrofonzugriff funktioniert in WSL i. d. R. nicht. Bitte normales Windows-Python nutzen.

py -m venv .venv

.\.venv\Scripts\activate

pip install -r requirements.txt

**Optional (GPU):**

set FASTER\_WHISPER\_DEVICE=cuda

set FASTER\_WHISPER\_COMPUTE\_TYPE=float16

CPU-Fallback (nichts setzen) nutzt automatisch eine sinnvolle Quantisierung.

**Start**

python main.py

* Klick auf **🎙 Aufnahme starten** → spricht ins Mikro
* Erneut klicken → Aufnahme stoppt, Transkription startet
* Ergebnis steht im großen Textfeld; Sprachkürzel erscheint in der Statuszeile

**Hinweise / Anpassungen**

* Modell wechseln: set FASTER\_WHISPER\_MODEL=large-v3-turbo (schneller)
* Empfohlene Sample-Rate ist 16 kHz; im Code konfigurierbar.
* Für Linux/Mac: Env-Vars entsprechend exportieren (export FASTER\_WHISPER\_DEVICE=cuda usw.).
* Wenn Torch/CUDA fehlt, läuft es automatisch auf CPU (langsamer).

Wenn du stattdessen ein **reines CLI** oder eine **Hotkey-Push-to-Talk**-Variante willst, sag Bescheid – ich gebe dir eine zweite Datei dazu.