

Benutzungsanleitung für Whisper OpenAI

Stand: Februar 2024 – Forschendes Lernen mit Oral History (FLOH)



Legende:

Anweisungen sind in: **rot und fett gefasst**

Eingabebestandteile sind zudem: *kursiv gesetzt*

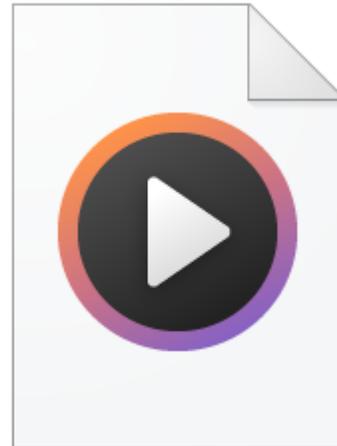
Ebenfalls sind Eingabebestandteile teils: farbig unterstrichen

Disclaimer:

Diese Anleitung erhebt keinen Anspruch auf absolute Vollständigkeit und garantiert kein reelles Gelingen des Vorhabens im jeweiligen Fall. Die Urheber lehnen jegliche Haftung für Schäden oder Folgeschäden, Verluste, entstandene Kosten oder Ausgaben, die mittelbar oder unmittelbar mit der Benutzung dieser Gebrauchsanleitung in Zusammenhang stehen, ab.

Vorbedingung für eine Transkription mit Whisper ist eine Audio-Datei.

→ Ein Oral History-Interview im Videoformat müsste bspw. im Vorhinein in eine *.flac*, *.wav* oder *.mp3*-Datei konvertiert werden.

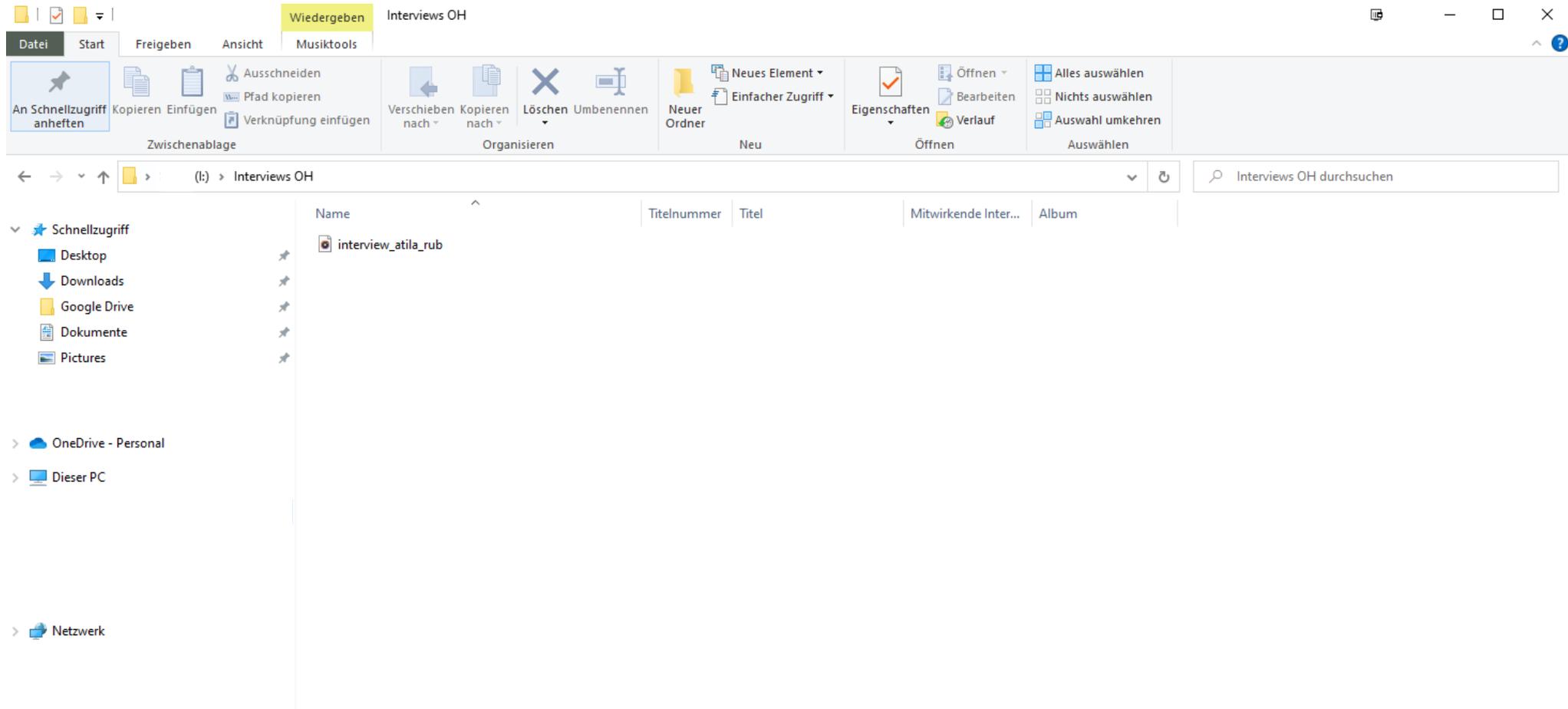


dateiname_beispiel.mp3|

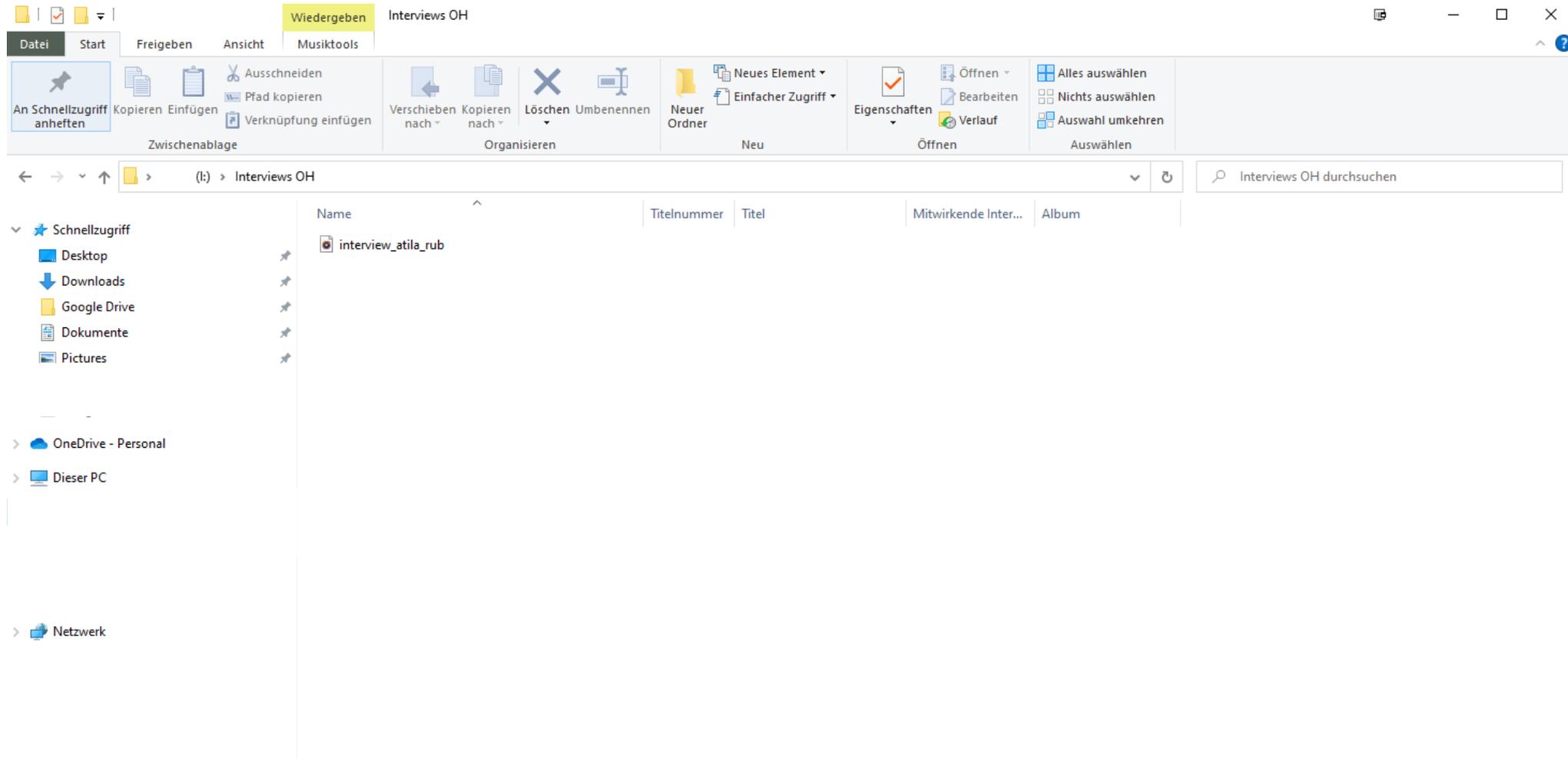
Die Audio-Datei sollte im Titel keine Leerzeichen und Sonderzeichen enthalten.
(Empfehlenswert sind Unterstriche)

Die zweite Vorbedingung ist ein extra Ordner zur Dateiablage.

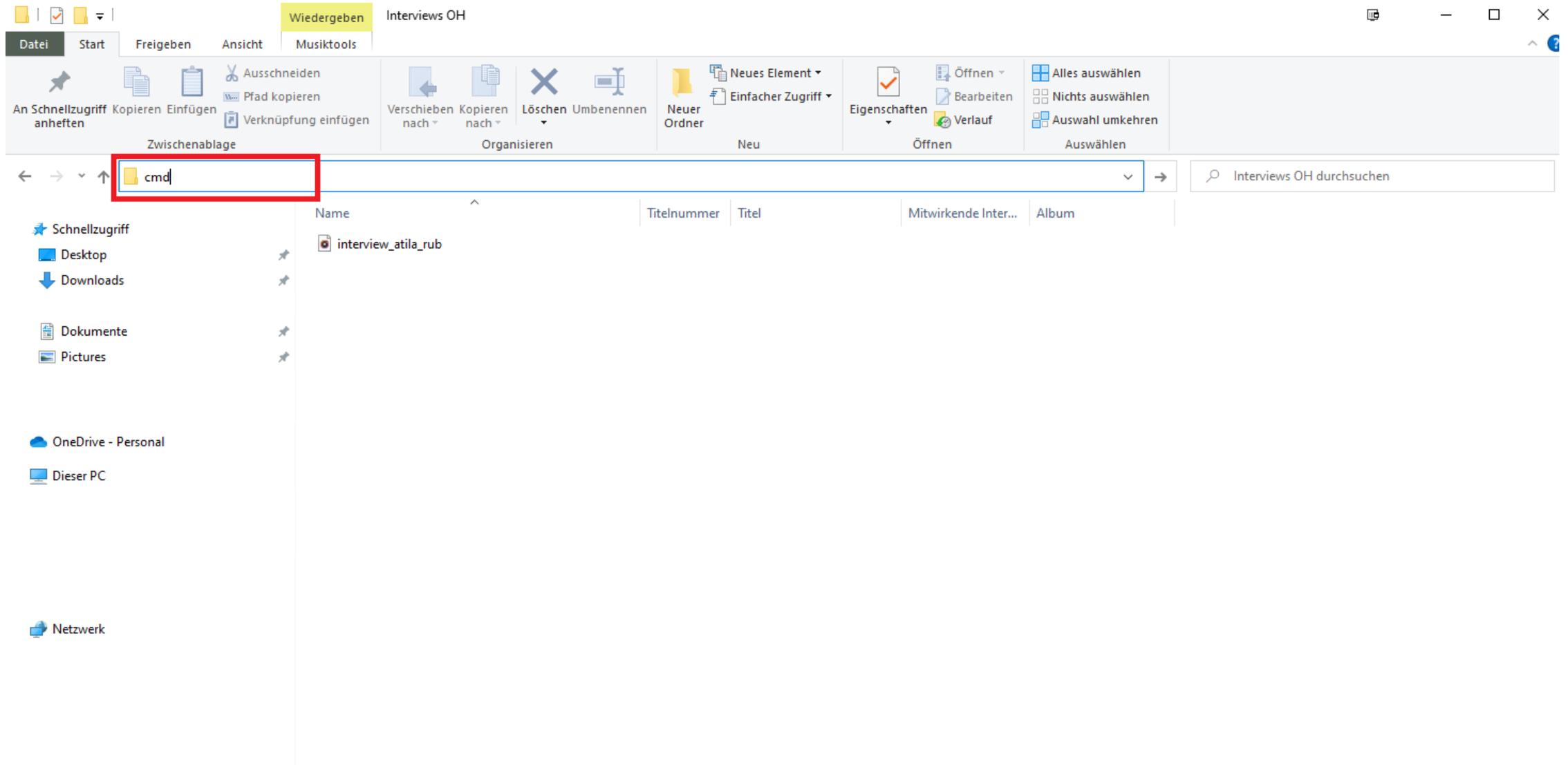
→ In diesen kopieren wir zu Anfang nun unsere Audio-Datei hinein.



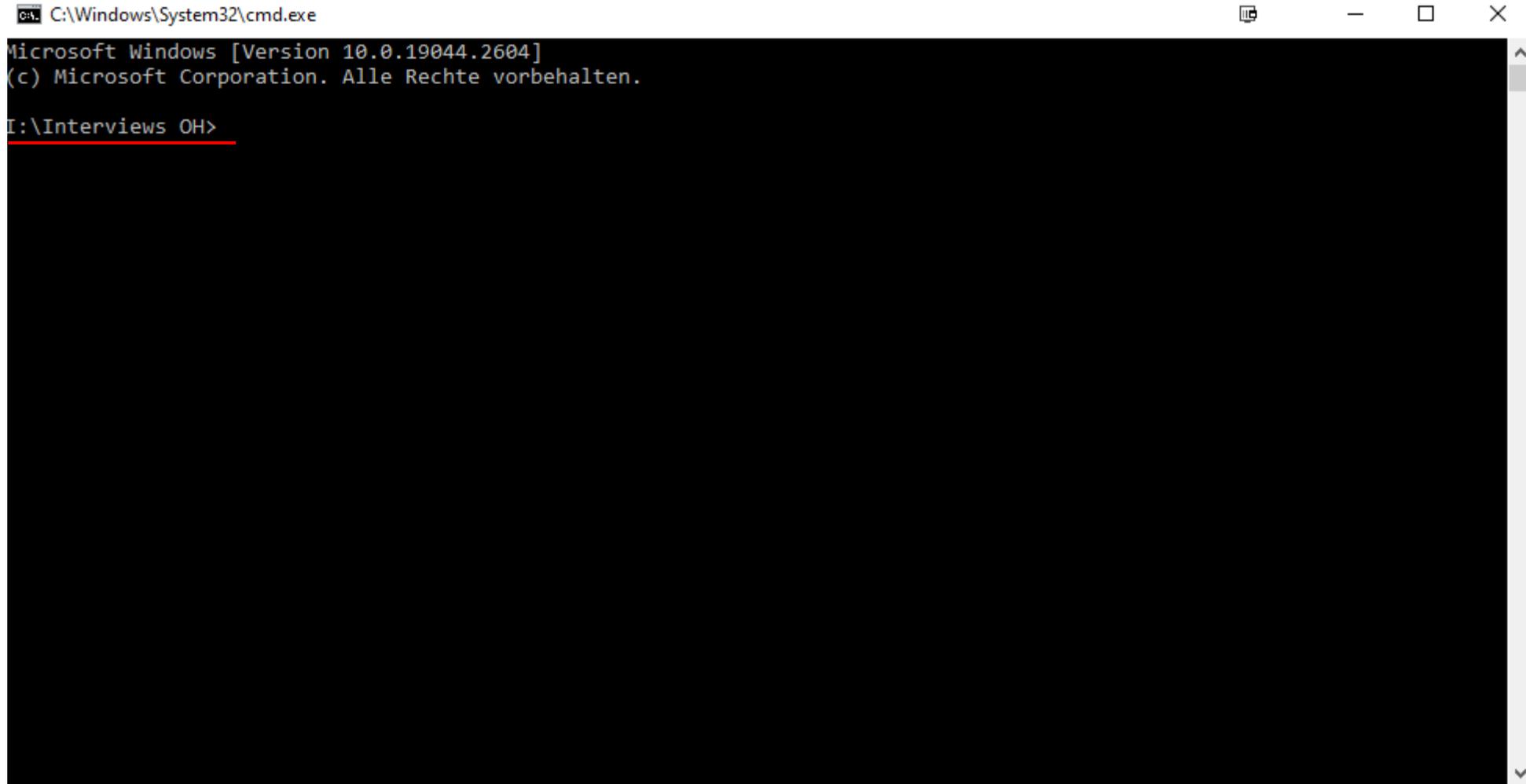
Die Transkription mit Whisper startet dann aus ebenjenem Ordner heraus.



Zuerst muss nun in den Dateipfad geklickt werden. Gib nun den Befehl „cmd“ ein und drücke die Taste „Enter“. Jetzt öffnet sich das Command-Fenster.



Das Command-Fenster ist nun aus dem Ordner heraus geöffnet, in dem die zu transkribierende Datei liegt. Bereit zum Start!



```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.2604]
(c) Microsoft Corporation. Alle Rechte vorbehalten.
I:\Interviews OH>
```

Nun muss eine Command-Line ausgewählt werden. Für Standard-Transkriptionen eignet sich der rot umrandete Befehl:

🔗 Command-line usage

The following command will transcribe speech in audio files, using the `medium` model:

```
whisper audio.flac audio.mp3 audio.wav --model medium
```

The default setting (which selects the `small` model) works well for transcribing English. To transcribe an audio file containing non-English speech, you can specify the language using the `--language` option:

```
whisper japanese.wav --language Japanese
```

Adding `--task translate` will translate the speech into English:

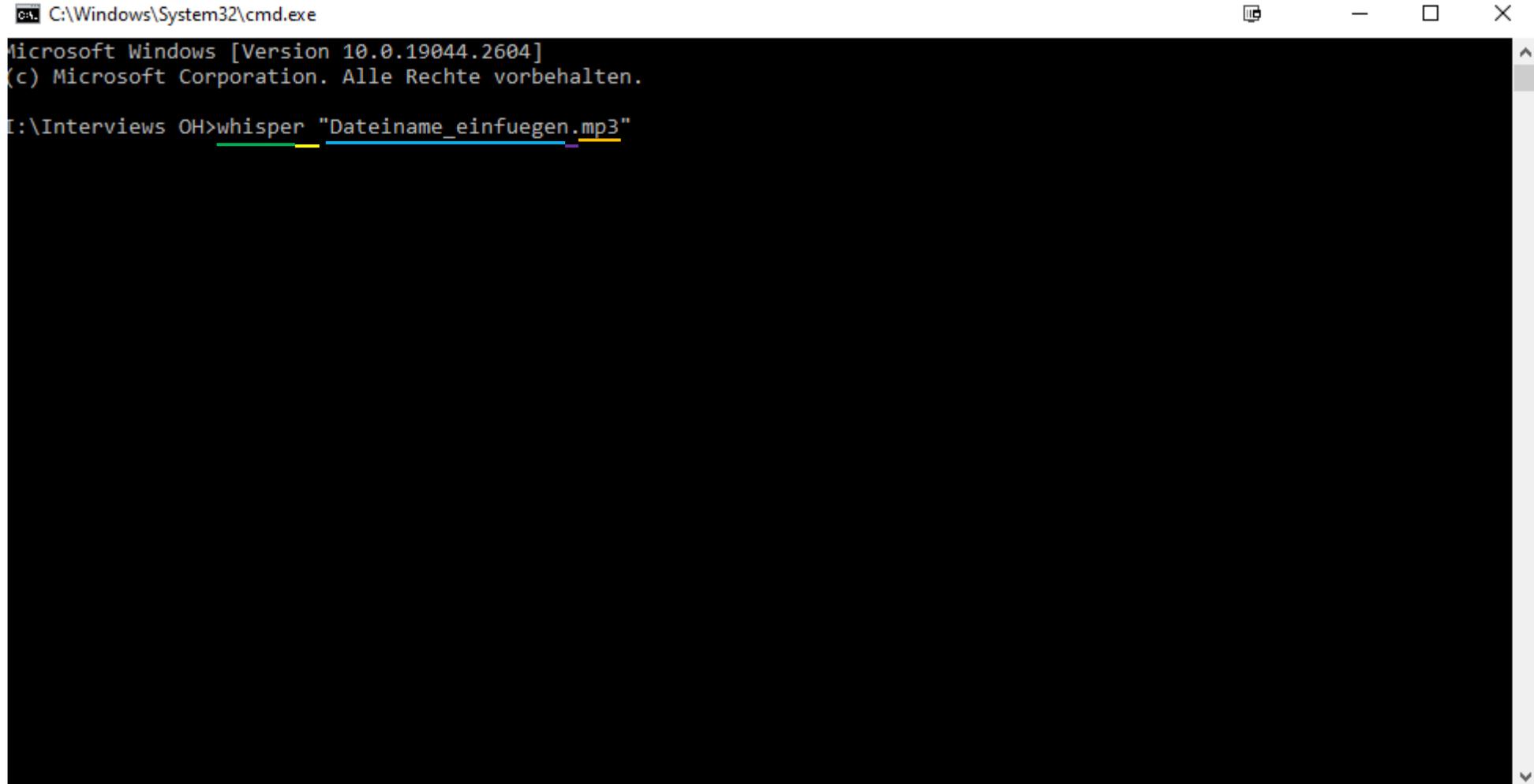
```
whisper japanese.wav --language Japanese --task translate
```

Run the following to view all available options:

```
whisper --help
```

See [tokenizer.py](https://github.com/openai/whisper/blob/main/tokenizer.py) for the list of all available languages.

Nun in die Zeile erst „whisper“, dann ein Leerzeichen und daraufhin in Anführungszeichen den Dateinamen eingeben. Wichtig!: Es muss zudem ein Punkt und daraufhin die Dateiendung der Audio-Datei eingegeben werden. (Bspw.: .mp3, .wav, etc.)

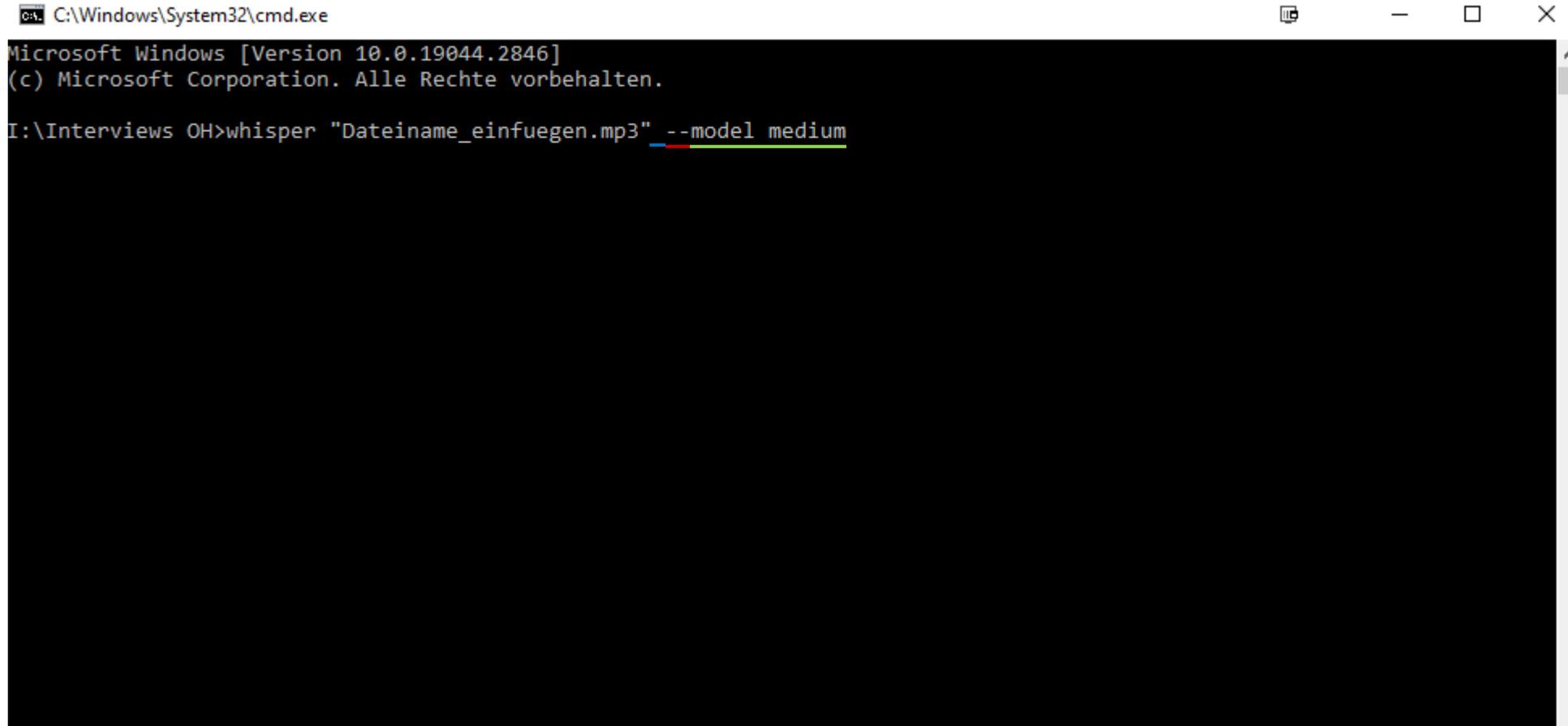


```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.2604]
(c) Microsoft Corporation. Alle Rechte vorbehalten.

I:\Interviews OH>whisper "Dateiname_einfuegen.mp3"
```

Daraufhin ein Leerzeichen und zwei Bindestriche eingeben. Dann „model“ und im Anschluss das Modell auswählen und den Namen eingeben (siehe folgende Übersicht!).

Mit der Taste „Enter“ startet die Transkription.



```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19044.2846]
(c) Microsoft Corporation. Alle Rechte vorbehalten.

I:\Interviews OH>whisper "Dateiname_einfuegen.mp3" --model medium
```

Übersicht der verfügbaren KI-Modelle und technische Anforderungen an den PC:

(Die Wahl sollte sich nach der technischen Leistungsfähigkeit des PCs, der verfügbaren Wartezeit und dem gewünschten Transkriptionserfolg richten. Das Medium-Modell war für uns meist die zeit- u. ressourceneffizienteste Option.)

→ Der Name ist jeweils auch der Befehl: bspw. „*medium*“

Available models and languages

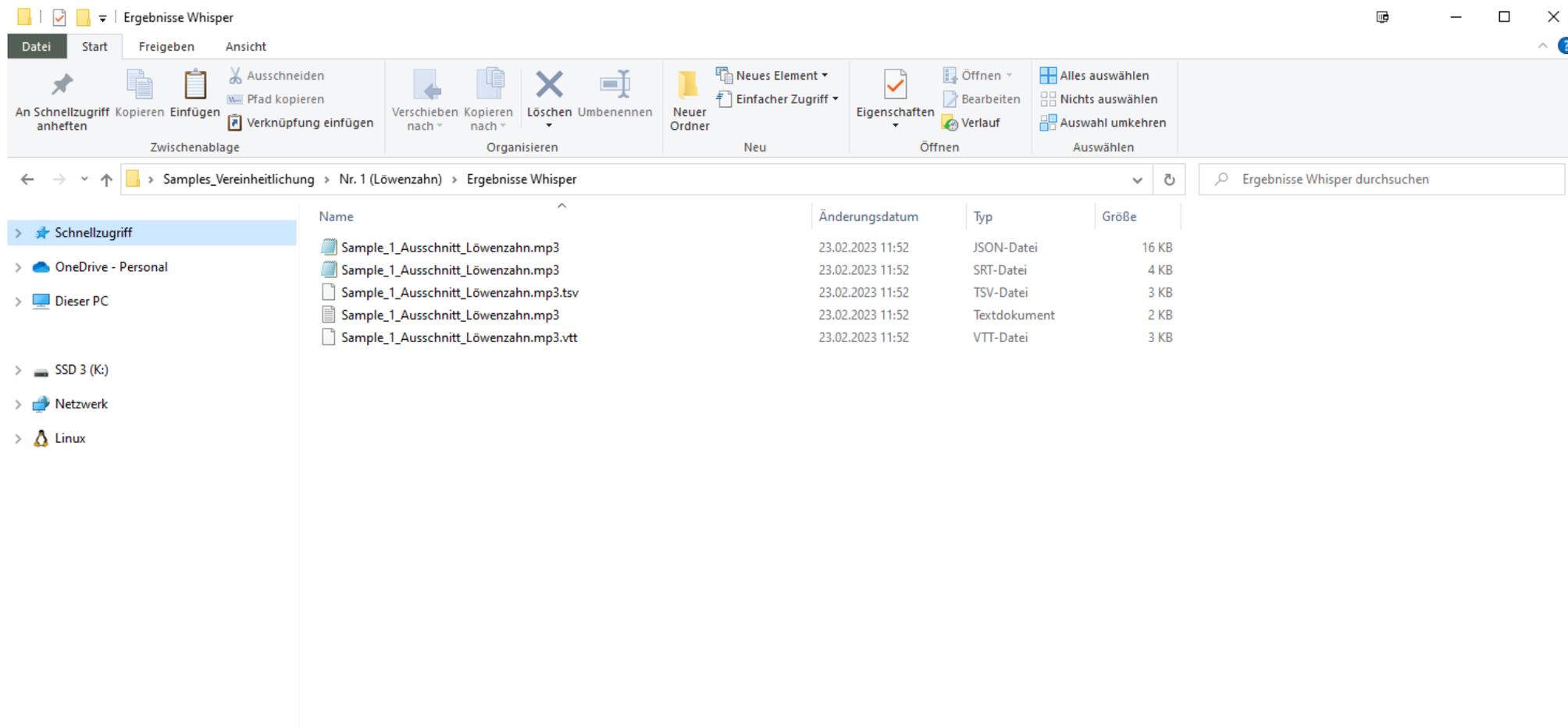
There are five model sizes, four with English-only versions, offering speed and accuracy tradeoffs. Below are the names of the available models and their approximate memory requirements and relative speed.

↓ *[erforderliche Leistung der Grafikkarte des PCs]*

Size	Parameters	English-only model	Multilingual model	Required VRAM	Relative speed
tiny	39 M	<code>tiny.en</code>	<code>tiny</code>	~1 GB	~32x
base	74 M	<code>base.en</code>	<code>base</code>	~1 GB	~16x
small	244 M	<code>small.en</code>	<code>small</code>	~2 GB	~6x
medium	769 M	<code>medium.en</code>	<code>medium</code>	~5 GB	~2x
large	1550 M	N/A	<code>large</code>	~10 GB	1x

For English-only applications, the `.en` models tend to perform better, especially for the `tiny.en` and `base.en` models. We observed that the difference becomes less significant for the `small.en` and `medium.en` models.

Der KI-Durchlauf endet automatisch. Das fertige Transkript speichert Whisper abschließend automatisiert in den Ordner der ursprünglichen Dateiablage. Das Transkript ist nun in fünf verschiedenen Dateiversionen weiterverwendbar.



Diese können nun bspw. im Text-Editor geöffnet oder für Untertitel von Video-Dateien verwendet werden.

The screenshot shows a Windows File Explorer window titled 'Ergebnisse Whisper'. The ribbon at the top includes 'Datei', 'Start', 'Freigeben', and 'Ansicht'. The main area displays a list of files in a table format:

Name	Änderungsdatum	Typ	Größe
Sample_1_Ausschnitt_Löwenzahn.mp3	23.02.2023 11:52	JSON-Datei	16 KB
Sample_1_Ausschnitt_Löwenzahn.mp3	23.02.2023 11:52	SRT-Datei	4 KB
Sample_1_Ausschnitt_Löwenzahn.mp3.tsv	23.02.2023 11:52	TSV-Datei	3 KB
Sample_1_Ausschnitt_Löwenzahn.mp3	23.02.2023 11:52	<u>Textdokument</u>	2 KB
Sample_1_Ausschnitt_Löwenzahn.mp3.vtt	23.02.2023 11:52	<u>VTT-Datei</u>	3 KB