

Transkription von Audio zu Text mit Hilfe von KI

Multilinguale Sprachverarbeitung durch das entwickelte Tool ScrAlbe

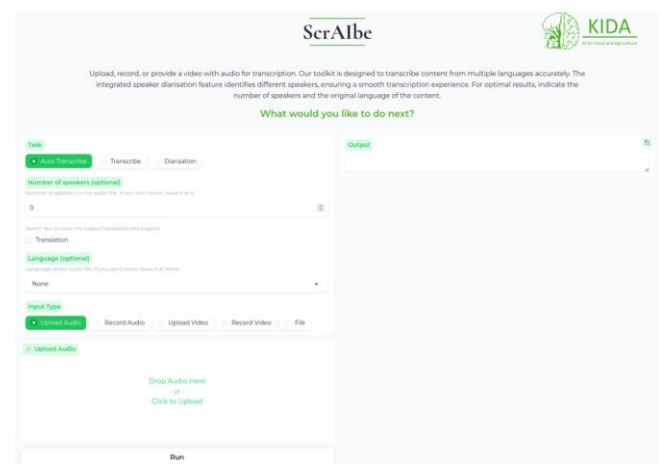
- ▶ **Hoher Bedarf an Verschriftlichung von Interviews, Vorträgen und Sitzungen etc.**
- ▶ **Präzise Transkription durch Kombination verschiedener KI-Modelle**
- ▶ **Benutzerfreundliche Oberfläche als auch Schnittstelle (API) verfügbar**

Hintergrund und Fragestellung

Im wissenschaftlichen und behördlichen Alltag entsteht sowohl durch Interviews, als auch durch Vorträge und Besprechungen viel **Audio- und Videomaterial**, das wissenschaftlich **ausgewertet** oder zu Protokollen **umgewandelt** werden soll. Bei der Verschriftlichung kann man sich mit einem **automatisierten Transkriptionstool** viel **Zeit** in der Erfassung **sparen** oder braucht keine teuren Transkriptionsdienste zu beauftragen. **Anschließend** können die Texte leicht mit anderen Anwendungen (wie dem KIDA AI Chat) **weiterverarbeitet** werden, um sie zu übersetzen oder Zusammenfassungen zu erstellen.

gesprochene Sprache in Text um.

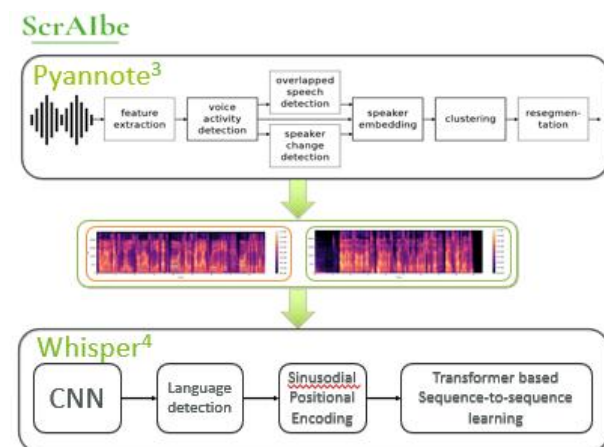
ScrAlbe-WebUI verbindet diese **beiden Funktionalitäten** mit einer benutzerfreundlichen Oberfläche unter Zuhilfenahme von **Gradio**:



Die Software kann auf einer eigenen Infrastruktur mit **benutzerfreundlicher Weboberfläche** betrieben werden. Zudem bietet sie für versiertere Nutzenden ein **Zugriff per Kommandozeile** und eine **Python API** an.

Ergebnisse und Schlussfolgerungen

ScrAlbe ermöglicht die **Identifikation verschiedener Sprechenden** und die **Transkription gesprochener Sprache** in Text in unterschiedlichen Sprachen durch die Nutzung verschiedener open-source Bibliotheken. Die Software **entlastet** die **Mitarbeitenden** im wissenschaftlichen und behördlichen Berufsalltag beim Verschriftlichen von Interviews, Vorträgen und Sitzungen. **Anschließend** können die Texte **leicht** mittels großer Sprachmodelle, wie sie der KIDA AI Chat verwendet, **weiterverarbeitet** werden. Durch die Möglichkeit der Verwendung der eigenen Infrastruktur, kann ScrAlbe **datenschutzkonform** betrieben werden ohne dass sensible Daten die eigene Einrichtung verlassen müssen.



Vorgehensweise

ScrAlbe verwendet zur Sprachverarbeitung die **open-source Bibliotheken** Whisper und Pyannote in Python:

- 1.) Mit Hilfe von **Pyannote**, das neuronale Netze verwendet, um die **Audioaufnahmen** nach SprecherInnen zu **segmentiert**, kann festgestellt werden, wer wann spricht. Dies ist besonders hilfreich in **Diskussionen und Besprechungen**, in denen es wichtig ist, wer welche Position vertreten hat.
- 2.) Anschließend wandelt **Whisper**, ein von OpenAI entwickeltes KI-Modell zur automatischen Spracherkennung, die

Kontakt	Informationen
KI-Beratung	KIDA-Bearbeitende: Jacob Schmieder (KIDA, DBFZ), Marco Henning (KIDA, Thünen), Aaron Roggenland (KIDA, MRI), Cristina Ortiz Cruz (KIDA, MRI)
Jacob Schmieder	Leitung KI-Beratung (KIDA): Micha Schneider (KIDA, Thünen) Software: https://github.com/kida4bmel/ScrAlbe