FREIE UNIVERSITÄT BERLIN

Oral-History.Digital

NFDI formemory

ASR4Memory

Ein KI-gestütztes Transkriptionsangebot für audiovisuelle Forschungsdaten

Dr. Tobias Kilgus, Peter Kompiel, Marc Altmann, Dr. Christian Horvat | Digitale Interview-Sammlungen | Universitätsbibliothek | Freie Universität Berlin

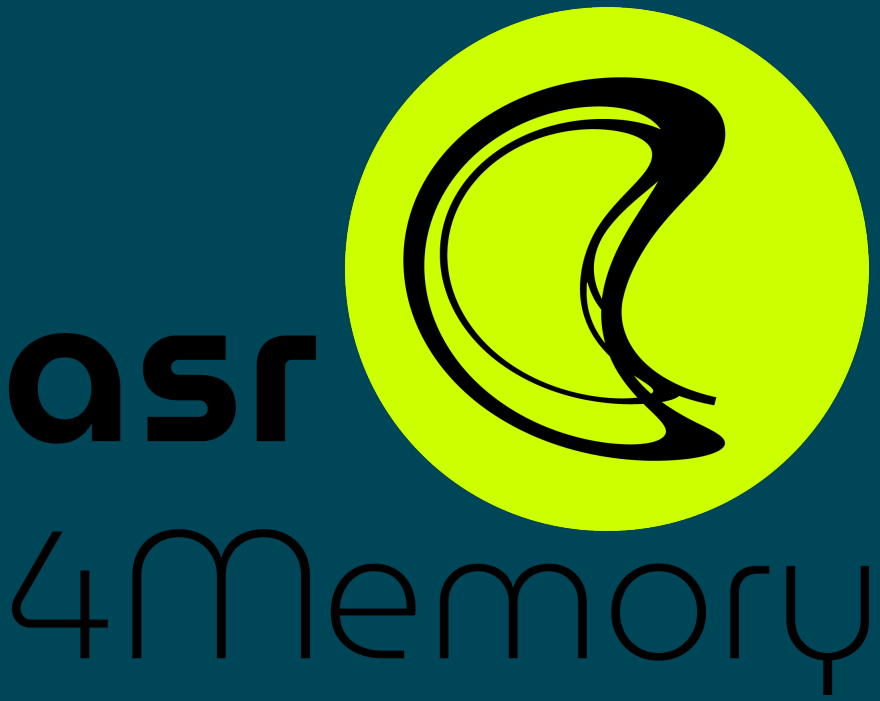
Projekt und Dienst

- **Team**
  - Tobias Kilgus, Peter Kompiel, Marc Altmann (alle FUB), Christian Horvat (FHNW, FB Mathematik)
- **Förderung**
  - NFDI, 4Memory, Incubator Funds 2024
- **Umsetzung**
  - Universitätsbibliothek der FU Berlin:
    - Abteilung Forschungs- & Publikationsservices
    - Team Digitale Interview-Sammlungen

Projekt und Dienst

- **NFDI4Memory-Serviceportfolio**
  - Seit 2025 als „4Memory-Initiativdienst“ gelistet
- **Zielgruppen**
  - Forschende der FUB, NFDI4Memory und aus anderen wissenschaftlichen Einrichtungen in DE/EU
  - Schwerpunkt auf den historisch arbeitenden Geisteswissenschaften (Oral History)
  - Einsetzbar in unterschiedlichen Fachdisziplinen/Szenarien
  - Keine Nutzungsgebühren

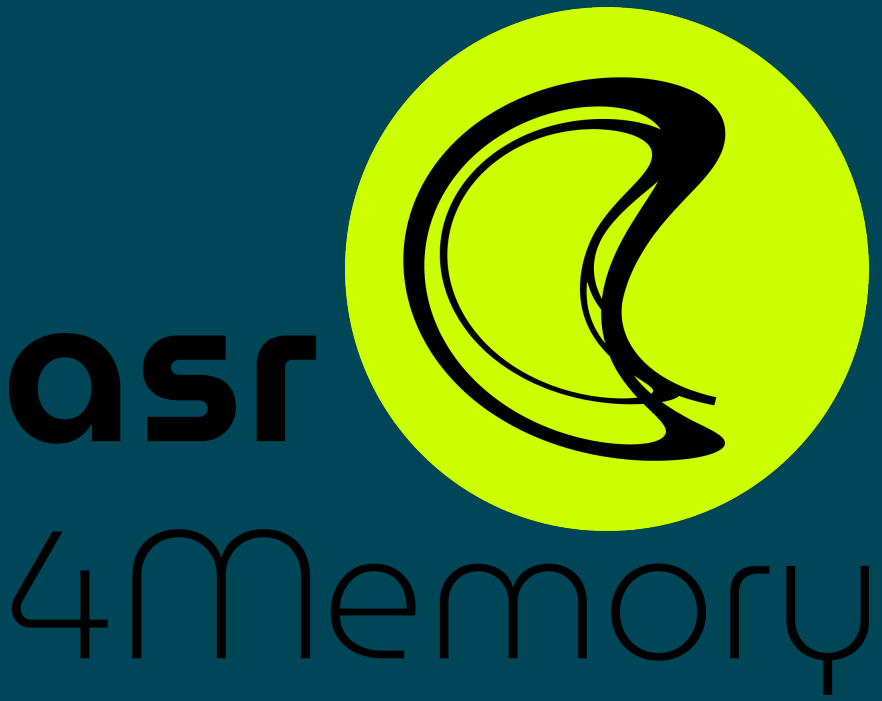
Warum haben wir das Projekt gemacht?



Ausgangslage

- **Sammlungen** von AV-Ressourcen liegen vor oder sind im Entstehen: Testimonials, Experteninterviews, Fachvorträge, etc.
  - → Inhaltlich/wissenschaftlich zu erschließen
  - → Grundlage ist Verschriftung
- **Bisheriges Vorgehen**:
  a) Manuelle Transkription, bspw. mit f4 oder Inqscribe
  ➔ Sehr zeitaufwändig, kostenintensiv, Sprachkompetenzen
  b) Kommerzielle Transkriptionsdienste
  ➔ (Z. T.) datenschutzproblematisch und kostenintensiv
  ➔ Ergebnisse oft nicht zufriedenstellend mit Blick auf Transkriptionsgenauigkeit, Timestamps und Formate

Wie sind wir vorgegangen?



Entwicklung

- **Bedarfsermittlung** in der 4Memory-Community
  → Austausch mit Sammlungsinhaber*innen, die mit audiovisuellen Forschungsdaten arbeiten:

  - Welche Audio-/Video-Quellen liegen vor?
  - Welchen Anforderungen und Schwierigkeiten bestehen?
  - Welche Transkriptformate werden benötigt?
  - Welche Infrastrukturen sind erforderlich?

Entwicklung

- **Nutzung der Automatischen Spracherkennung (ASR)**:
  → Umwandlung gesprochener Sprache in Text

In modern times, we expect more of our automatic systems. The task of **automatic speech recognition** (**ASR**) is to map any waveform like this:

to the appropriate string of words:

```
It's time for lunch!
```

(Jurafsky & Martin, 2023)

Entwicklung

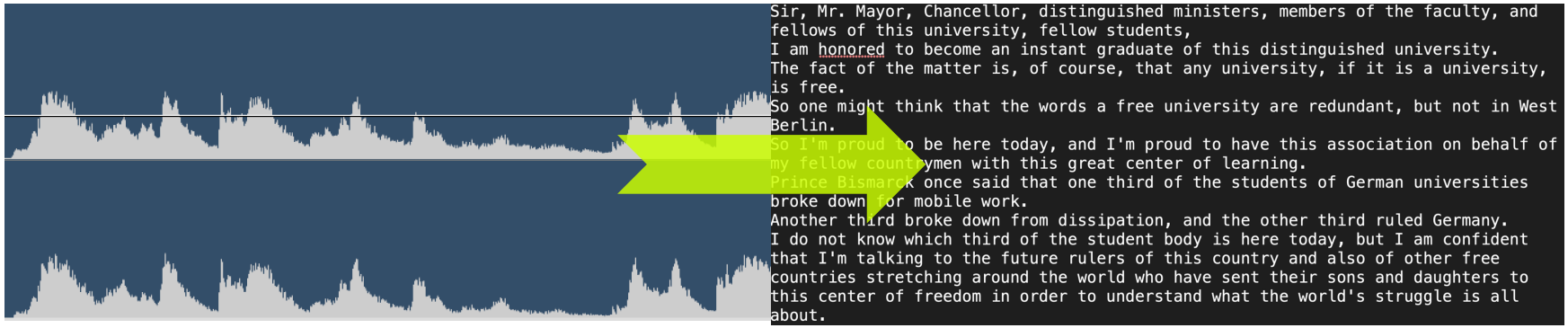
- Einsatz von **Künstlicher Intelligenz** (KI) bei der **ASR**
  - → Evaluation verschiedener ASR-Modelle
  - → **„WhisperX“:** Integration der „Whisper“-Reimplementierung (Universität Oxford), entwickelt von OpenAI (ChatGPT)

- **Open-Source-Code und -Gewichte**
  - ➔ Quelloffene Lizenz (BSD-2): Code flexibel und anpassbar
  - ➔ Fehlende Trainingsdaten: faktisch Open-Weights-Modell
  - ➔ 1,55 Mrd. Parameter: hocheffizientes Transformer-Modell

Entwicklung

- **Option 1: Web-Service und Transkriptions-Pipeline mit dem Media Management Tool (MMT)**

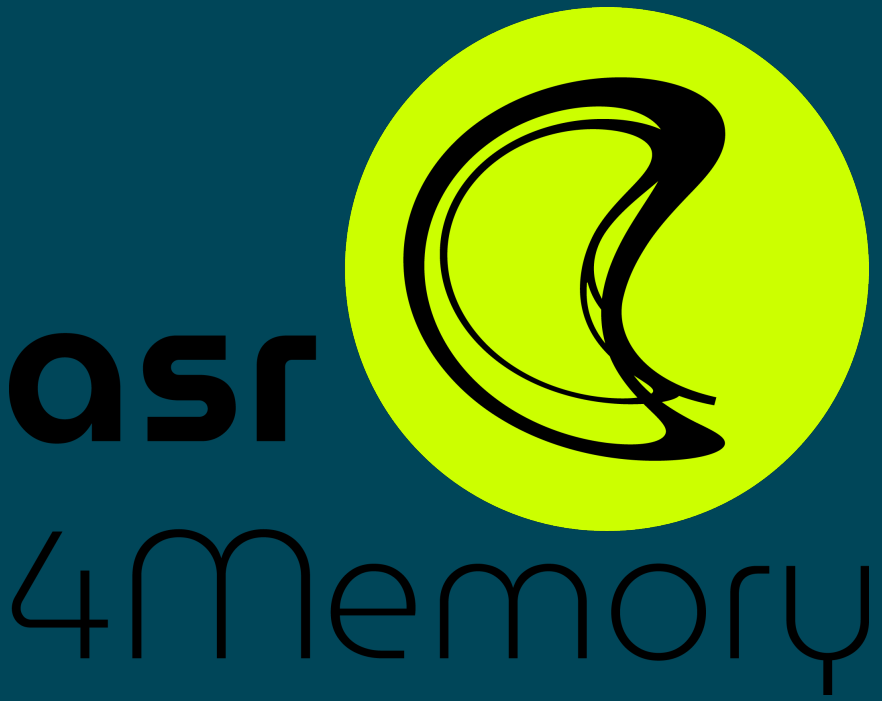
Upload der AV-Daten via MMT → Prüfung und Optimierung der Tonspur → Automatische Transkription (ASR) → Generierung der Exportformate → Bereitstellung der Formate via MMT



Entwicklung

- **Option 2: Nutzung des Open-Source-Codes**
  - Ermöglicht lokale Eigeninstallation der Software
  - Weiterentwicklung und -verteilung in der Community
  - Open-Source-Lizenz: AGPL-3.0
- **Github-Repositorien**: https://github.com/asr4memory
  - Transkription: https://github.com/asr4memory/asr-transcribe
  - Evaluation: https://github.com/asr4memory/asr-evaluate
  - Tonoptimierung: https://github.com/asr4memory/asr-optimize
  - Upload-Tool: https://github.com/asr4memory/mmt-py

Welche Ergebnisse
wurden erreicht?



Ergebnisse

- **Web-Service:**

  - **Bereitstellung** der AV-Daten zur Transkription über das **Media Management Tool (MMT)**
  - **Browserbasierte Anwendung**, keine Installation notwendig
  - **Zugang**: https://mmt.oral-history.digital/

Ergebnisse

- **Datenschutzkonform**:

  - **Web-Service**: Datenverarbeitung ausschließlich in der IT-Infrastruktur der FUB
  - **Sicheres Datenmanagement** auf Forschungsspeicher:
    - Prüfsummen
    - Backups
    - Virenscan
    - Single-Sign-On (SSO) mit MFA (OAuth-Standard)
    - Sichere Verwaltung der hochgeladenen Daten in MMT: Nur Sie haben Zugriff auf Ihre Daten.

Ergebnisse

- **Performance**:
  - Schnelle Verarbeitung (RTF = ~0,01 → 4h in 2 min)
  - Hochwertige Transkription (niedrige WER), wenn:
    - Sprache ausreichend in Trainingsdaten vertreten
    - Klares Sprachsignal, wenig Akzent/Dialekt

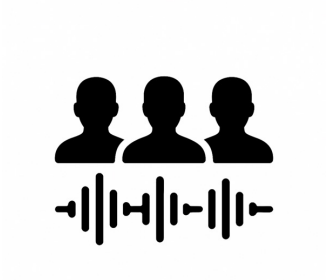
- **Open Source:**
  - Installation und Betrieb auf eigenem Rechner
    → Keine Clouds, externe Dienste, Datenabflüsse

Ergebnisse

- **Mehrsprachigkeit**:

  - Unterstützung von etwa 30 Sprachen: deutsch, englisch, spanisch, ukrainisch, usw. (Gesamtliste)
  - Automatische Sprachdetektion möglich

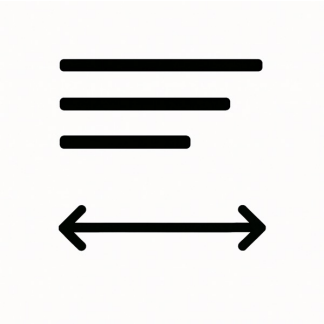
- **Diarisierung:**



  - Akustische Erkennung und Annotation der Sprecher*innen
  - Sprechpausen anzeigen, z. B. Annotation einer zweisekündigen Pause: <p2> oder [2 seconds]

Ergebnisse

- **Segmentlänge:**



  - Intelligente, dynamische Begrenzung der Zeichen pro Segment

- **Optimierung der Tonspur:**

  - Für die optimale Spracherkennung und Reduktion von Halluzinationen:
    → Filterung, Normalisierung, Konvertierung

Ergebnisse

- **Alignierung:**

  - Zeitkodierung mit Millisekunden-genauen Zeitmarken
  - Synchronisierung von Transkript mit AV-Medium
  - Verbesserte Auffindbarkeit und Nachnutzbarkeit
  - Weitere Nutzungsmöglichkeiten, u. a. in der Textanalyse (NER, TM), Anonymisierung

1	WORD	START	END
2	Ja.	00:00:02.077	00:00:02.398
3	Darf	00:00:02.438	00:00:02.637
4	ich	00:00:02.677	00:00:02.758
5	rein?	00:00:02.959	00:00:03.738

Ergebnisse

- **Exportformate:**

  - Export der Transkripte in verschiedene Formate
  - Bereitstellung des Exports im standardisierten, plattformunabhängigen BagIt-Format

Ergebnisse

- **Exportformate**: TXT, PDF/A | VTT, SRT

```
Sir, Mr. Mayor, Chancellor, distinguished ministers, members of
the faculty, and fellows of this university, fellow students,
I am honored to become an instant graduate of this distinguished
university.
The fact of the matter is, of course, that any university, if it
is a university, is free.
So one might think that the words a free university are
redundant, but not in West Berlin.
So I'm proud to be here today, and I'm proud to have this
association on behalf of my fellow countrymen with this great
center of learning.
Prince Bismarck once said that one third of the students of
German universities broke down for mobile work.
Another third broke down from dissipation, and the other third
ruled Germany.
I do not know which third of the student body is here today, but
I am confident that I'm talking to the future rulers of this
country and also of other free countries stretching around the
world who have sent their sons and daughters to this center of
```

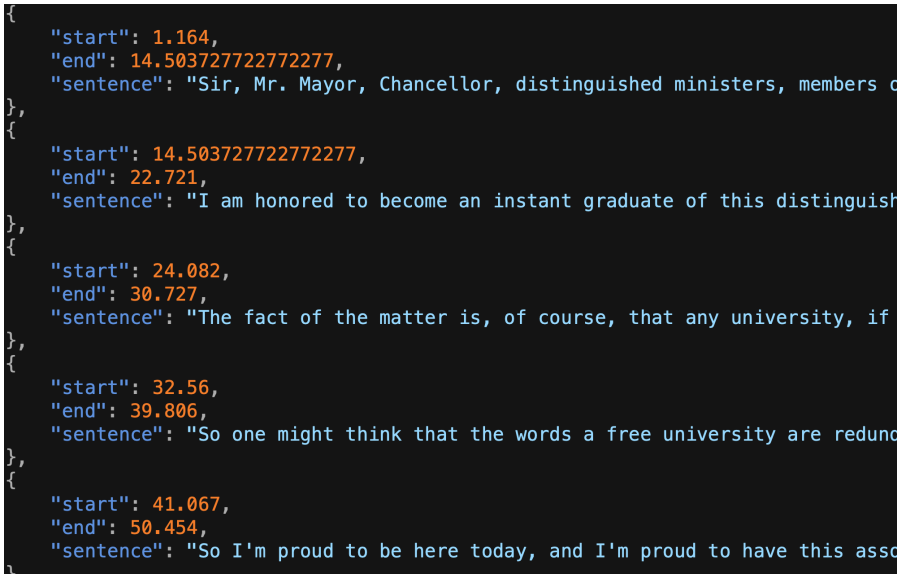
```
WEBVTT

1
00:00:01.164 --> 00:00:14.503
Sir, Mr. Mayor, Chancellor, distinguished ministers, members of the
faculty, and fellows of this university, fellow students,

2
00:00:14.503 --> 00:00:22.721
I am honored to become an instant graduate of this distinguished
university.

3
00:00:24.082 --> 00:00:30.727
The fact of the matter is, of course, that any university, if it is a
university, is free.

4
00:00:32.560 --> 00:00:39.806
So one might think that the words a free university are redundant,
```

→ Zur manuellen Nachbearbeitung (z. B. in MAXQDA) und zur Langzeitarchivierung

| → Untertitelung von AV-Medien in Playern (Barrierefreiheit)

Ergebnisse

- **Exportformate**: JSON, TEI-XML | CSV



```
{
    "start": 1.164,
    "end": 14.503727722772277,
    "sentence": "Sir, Mr. Mayor, Chancellor, distinguished ministers, members of
},
{
    "start": 14.503727722772277,
    "end": 22.721,
    "sentence": "I am honored to become an instant graduate of this distinguished
},
{
    "start": 24.082,
    "end": 30.727,
    "sentence": "The fact of the matter is, of course, that any university, if it
},
{
    "start": 32.56,
    "end": 39.806,
    "sentence": "So one might think that the words a free university are redundar
},
{
    "start": 41.067,
    "end": 50.454,
    "sentence": "So I'm proud to be here today, and I'm proud to have this associ
```

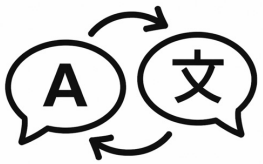
IN	TRANSCRIPT
00:00:01.164	Sir, Mr. Mayor, Chancellor, distinguished ministers, members of the faculty, and fellows of this university, fe
00:00:14.504	I am honored to become an instant graduate of this distinguished university.
00:00:24.082	The fact of the matter is, of course, that any university, if it is a university, is free.
00:00:32.560	So one might think that the words a free university are redundant, but not in West Berlin.
00:00:41.067	So I'm proud to be here today, and I'm proud to have this association on behalf of my fellow countrymen v
00:00:52.256	Prince Bismarck once said that one third of the students of German universities broke down for mobile wo
00:01:01.570	Another third broke down from dissipation, and the other third ruled Germany.
00:01:08.713	I do not know which third of the student body is here today, but I am confident that I'm talking to the future
00:01:35.037	I know that when you leave this school, you will not imagine that this institution was founded by citizens of
00:01:48.189	and was developed by citizens of West Berlin, that you will not imagine that these men who teach you hav
00:02:10.448	This school is not interested in turning out merely corporation lawyers or skilled accountants.
00:02:18.774	What it is interested in, and this must be true of every university, It must be interested in turning out citizen
00:02:28.925	men who comprehend the difficult, sensitive tasks that lie before us as free men and women, and men wh
00:02:44.141	That's why you're here, and that's why this school was founded, and all of us benefit from it.
00:03:08.126	It is a fact that in my own country, in the American Revolution, that revolution and the society developed th

→ Zur automat. Datenverarbeitung in Systemen (z. B. Information Retrieval)

→ Zukünftige Formate: IIIF-AV, CMDI, EAD etc. ?

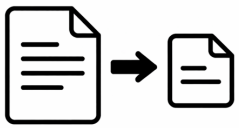
Ergebnisse

- **Übersetzung**:

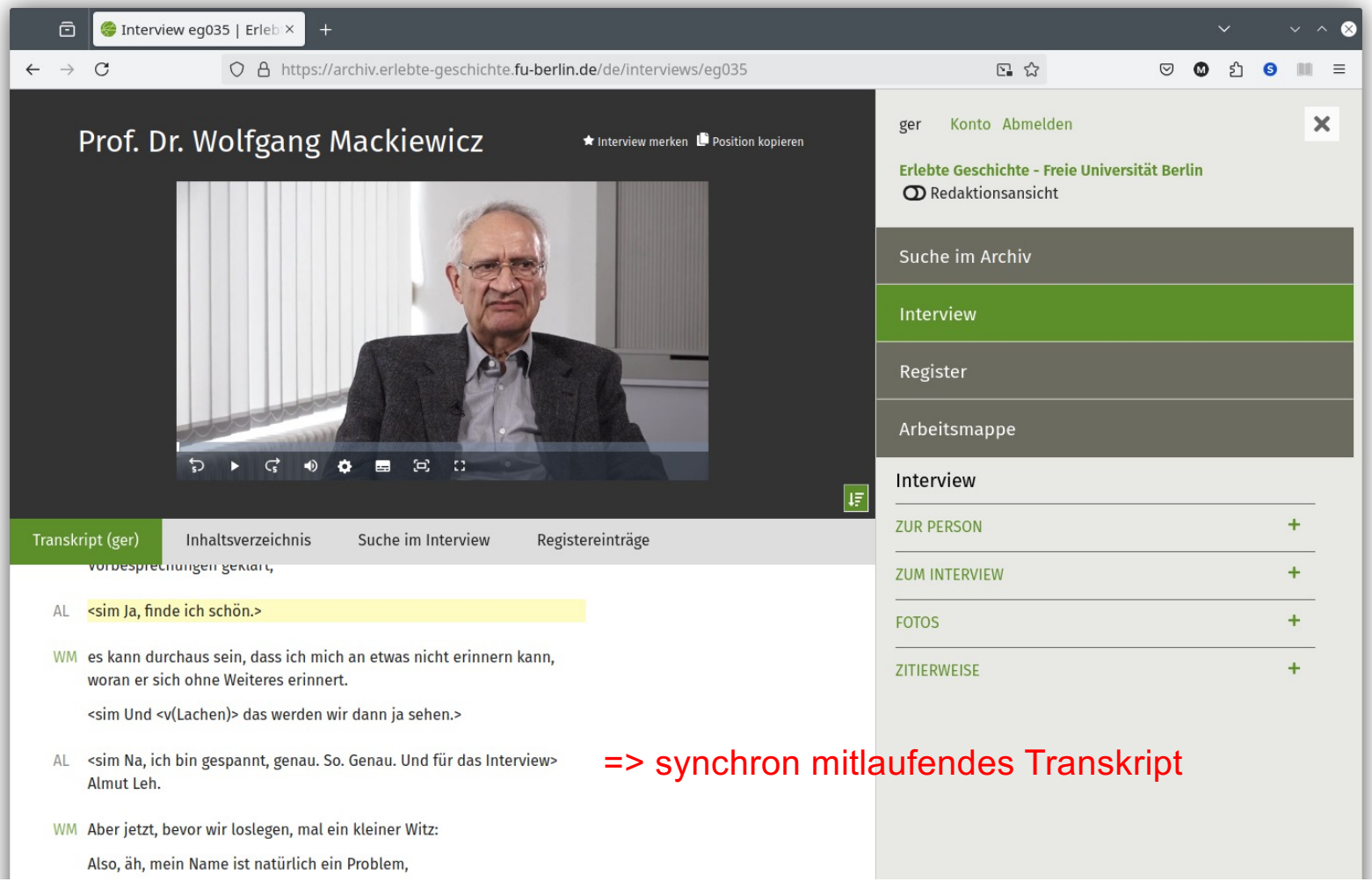


  - Die Transkripte werden ins Englische übersetzt (perspektivisch in weitere Sprachen).

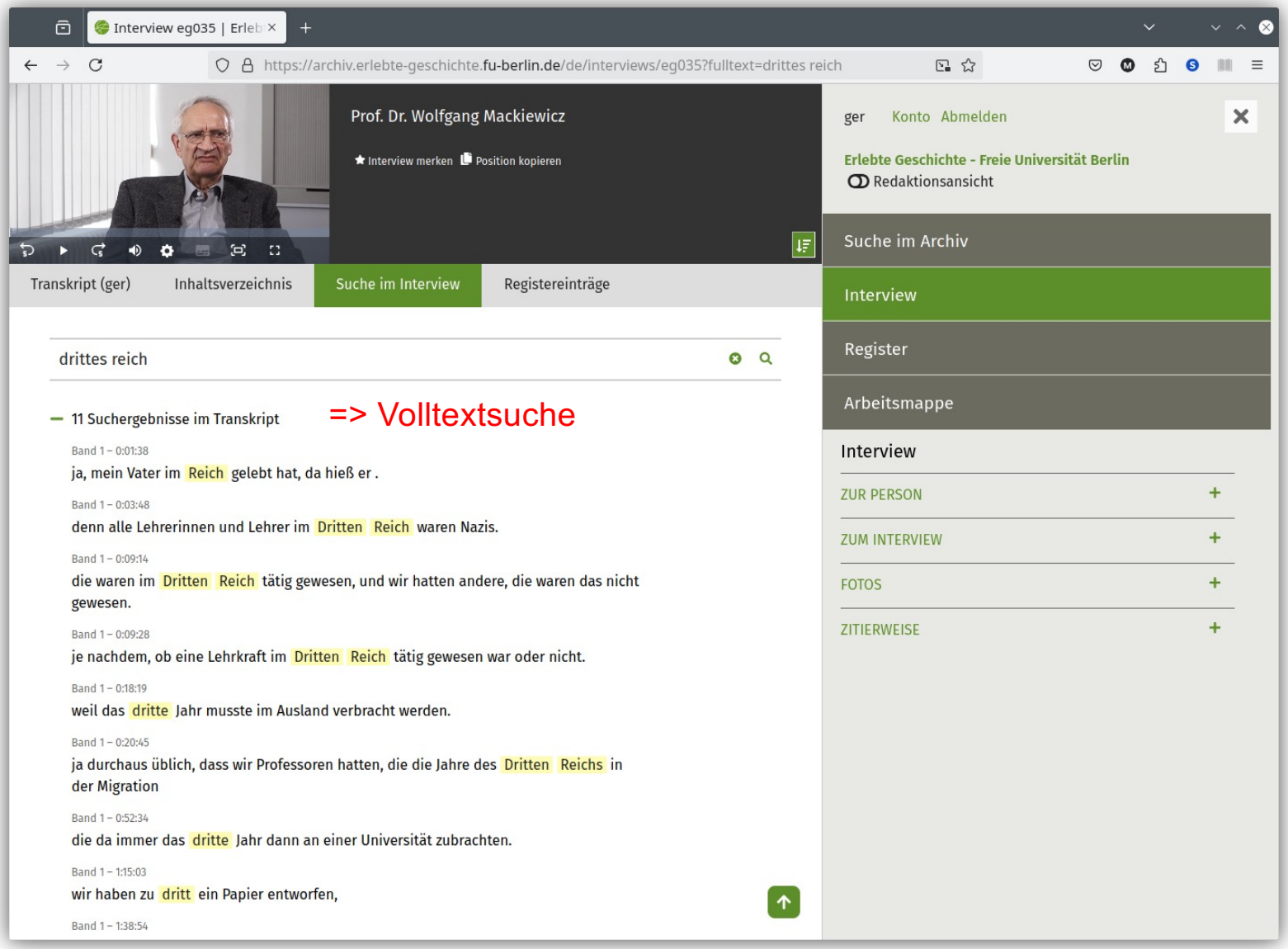
- **Zusammenfassungen**:

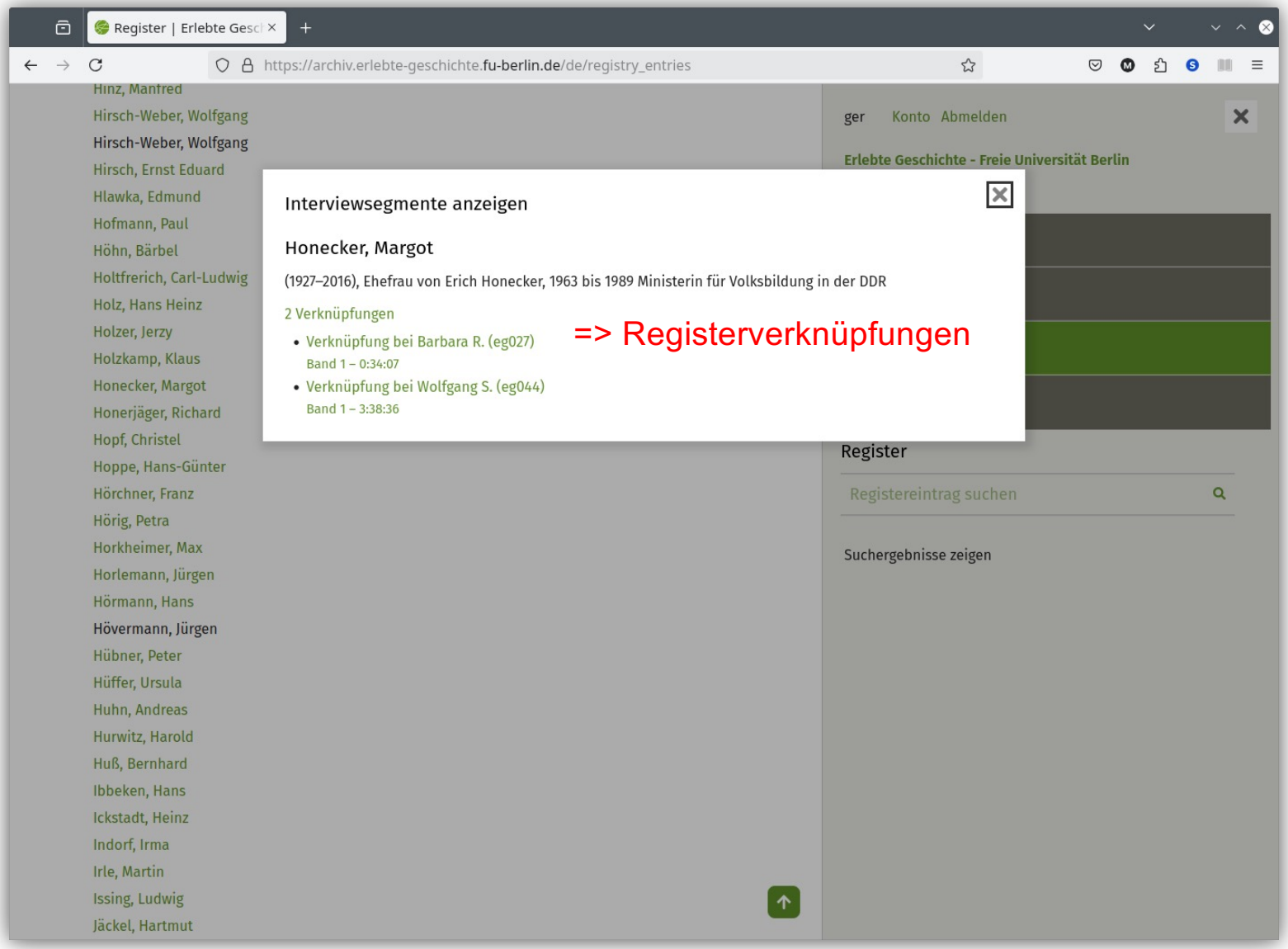


  - Aus dem Transkript werden mehrsprachige Abstracts mithilfe eines lokal ausgeführten Sprachmodells erstellt.



=> synchron mitlaufendes Transkript

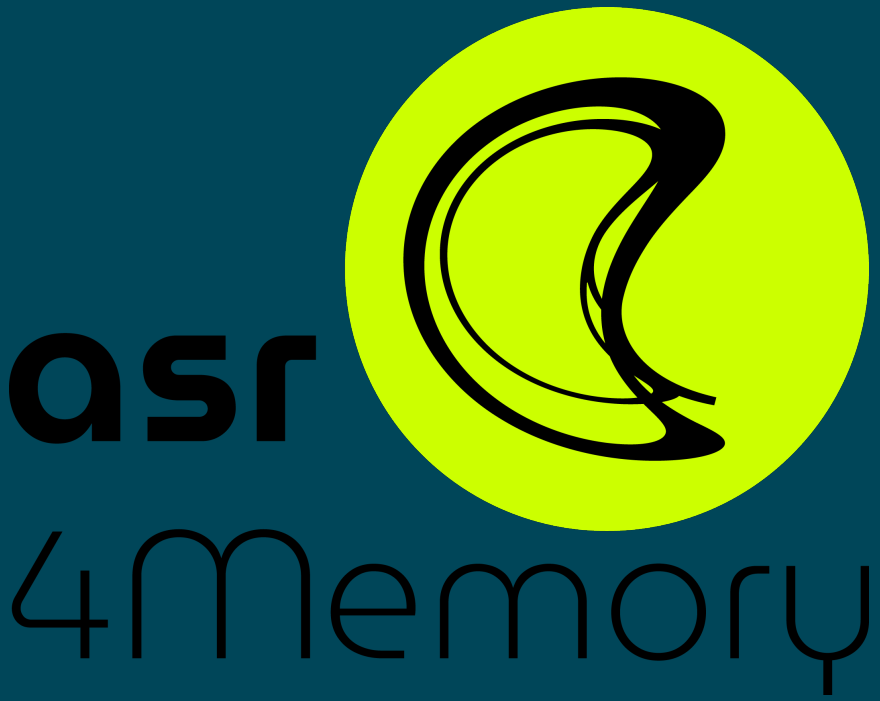

Interview eg035 | Erleb
https://archiv.erlebte-geschichte.fu-berlin.de/de/interviews/eg035?fulltext=drittes reich
Prof. Dr. Wolfgang Mackiewicz
Interview merken
Position kopieren
ger
Konto
Abmelden
Erlebte Geschichte - Freie Universität Berlin
Redaktionsansicht
Suche im Archiv
Interview
Register
Arbeitsmappe
Transkript (ger)
Inhaltsverzeichnis
Suche im Interview
Registereinträge
drittes reich
11 Suchergebnisse im Transkript
=> Volltextsuche
Band 1 – 0:01:38
ja, mein Vater im Reich gelebt hat, da hieß er .
Band 1 – 0:03:48
denn alle Lehrerinnen und Lehrer im Dritten Reich waren Nazis.
Band 1 – 0:09:14
die waren im Dritten Reich tätig gewesen, und wir hatten andere, die waren das nicht gewesen.
Band 1 – 0:09:28
je nachdem, ob eine Lehrkraft im Dritten Reich tätig gewesen war oder nicht.
Band 1 – 0:18:19
weil das dritte Jahr musste im Ausland verbracht werden.
Band 1 – 0:20:45
ja durchaus üblich, dass wir Professoren hatten, die die Jahre des Dritten Reichs in der Migration
Band 1 – 0:52:34
die da immer das dritte Jahr dann an einer Universität zubrachten.
Band 1 – 1:15:03
wir haben zu dritt ein Papier entworfen,
Band 1 – 1:38:54
Interview
ZUR PERSON
ZUM INTERVIEW
FOTOS
ZITIERWEISE


Register | Erlebte Gesc
https://archiv.erlebte-geschichte.fu-berlin.de/de/registry_entries
Hinz, Manfred
Hirsch-Weber, Wolfgang
Hirsch-Weber, Wolfgang
Hirsch, Ernst Eduard
Hlawka, Edmund
Hofmann, Paul
Höhn, Bärbel
Holtfrerich, Carl-Ludwig
Holz, Hans Heinz
Holzer, Jerzy
Holzkamp, Klaus
Honecker, Margot
Honerjäger, Richard
Hopf, Christel
Hoppe, Hans-Günter
Hörchner, Franz
Hörig, Petra
Horkheimer, Max
Horlemann, Jürgen
Hörmann, Hans
Hövermann, Jürgen
Hübner, Peter
Hüffer, Ursula
Huhn, Andreas
Hurwitz, Harold
Huß, Bernhard
Ibbeken, Hans
Ickstadt, Heinz
Indorf, Irma
Irle, Martin
Issing, Ludwig
Jäckel, Hartmut
Interviewsegmente anzeigen
Honecker, Margot
(1927–2016), Ehefrau von Erich Honecker, 1963 bis 1989 Ministerin für Volksbildung in der DDR
2 Verknüpfungen
Verknüpfung bei Barbara R. (eg027)
Band 1 – 0:34:07
Verknüpfung bei Wolfgang S. (eg044)
Band 1 – 3:38:36
=> Registerverknüpfungen
ger
Konto
Abmelden
Erlebte Geschichte - Freie Universität Berlin
Register
Registereintrag suchen
Suchergebnisse zeigen

Ergebnisse

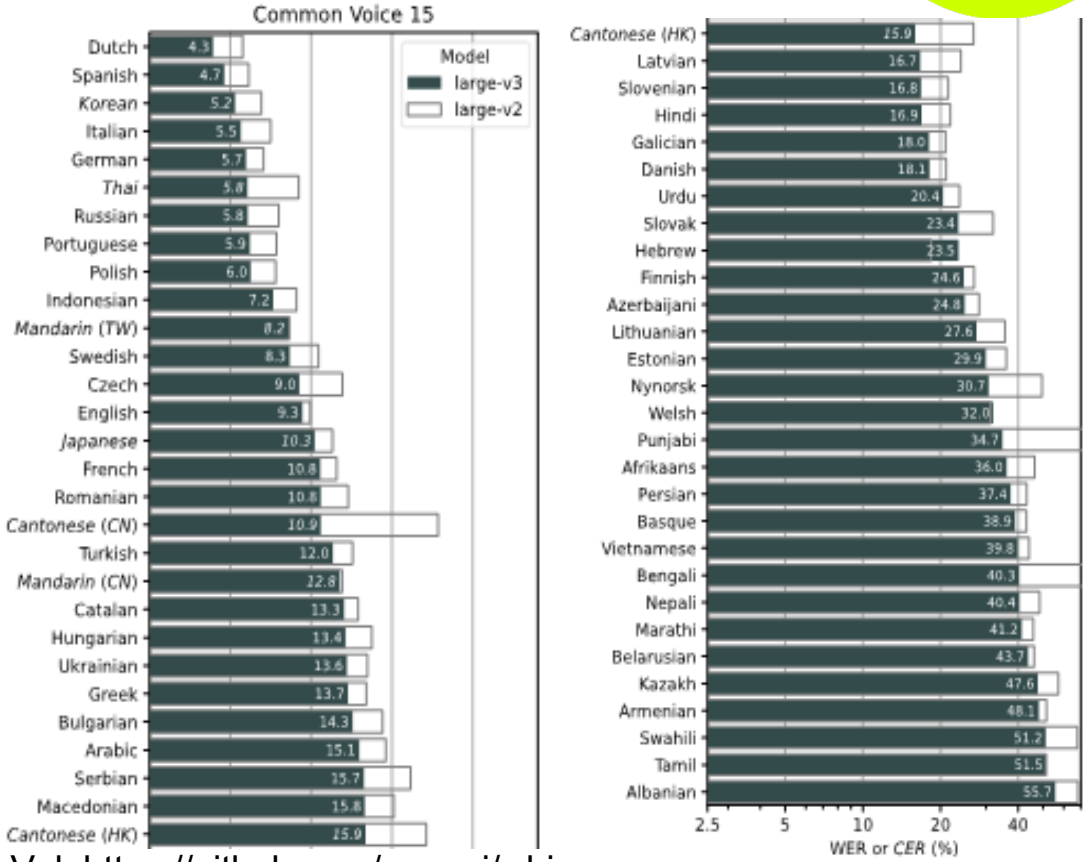
- **über 50 Einrichtungen** (Universitäten, Archive, Gedenkstätten, Bibliotheken) haben den Dienst ASR4Memory bis dato genutzt
  - z. B. Technische Universität Berlin, FernUniversität in Hagen, Universität Hamburg, Österreichische Mediathek, Staatliche Archive Bayerns, Gedenkstätten Gestapokeller und Augustaschacht, Stiftung Flucht, Vertreibung und Versöhnung, Bundeskanzler-Helmut-Schmidt-Stiftung, Forschungsstelle für Zeitgeschichte in Hamburg, Europa-Universität Flensburg, Friedrich-Alexander-Universität Erlangen, usw.
  - FUB: u. a. Universitätsarchiv, Charité, FB WiWiss, Friedrich-Meinecke-Institut, Kommunikation und Marketing

Welche Schwächen hat die ASR – und wie adressieren wir diese?



Schwächen der ASR

- **Limitierte Sprachen-Unterstützung**:
  - Für das Training von Whisper wurden 680.000 Stunden Audio-material mit z.T. Referenztrans-kripten aus dem Internet verwen-det, ~ 100 Sprachen, überwiegend englisch → Bias!
  - Je stärker Sprache in Datensatz vertreten, desto besser die ASR (Ausnahmen bestätigen die Regel)



Vgl. https://github.com/openai/whisper

Schwächen der ASR

1) Füllwörter

OT: „Der hat uns da, äh äh, ein paar Sachen, äh äh, Schriftstücke gezeigt, die also eigentlich vollkommen belanglos waren.“[1]

WhisperX: „Der hat uns da ein paar Sachen, Schriftstücke gezeigt, die also eigentlich vollkommen belanglos waren.“

[1] Schachtschneider, Olaf , Interview ev001, 03.06.2019, Band 1 – 0:06:07, Interview-Archiv "Eiserner Vorhang", https://archiv.eiserner-vorhang.de/de/interviews/ev001?tape=1&time=0h06m07s, 07.03.2024

Schwächen der ASR

2) Wiederholungen / Satzabbrüche / Verzögerungen

OT: „Wir wohnten damals in Berlin-Köpenick. In_ in_. Äh, na, eine Hütte war es ja eigentlich gewesen.“[2]

WhisperX: „Wir wohnten damals in Berlin-Köpenick. In einer Hütte war es ja eigentlich.“

[2] Schachtschneider, Olaf , Interview ev001, 03.06.2019, Band 1 – 0:01:09, Interview-Archiv "Eiserner Vorhang", https://archiv.eiserner-vorhang.de/de/interviews/ev001?tape=1&time=0h01m09s, 07.03.2024

Schwächen der ASR

3) Dialekte

„ick“ → „ich“, „jewesen“ → „gewesen“, „kriech‘ ta“ → „kriegte er“ usw.

OT: „Nach einer Woche oder so als Säugling, 14 Tage später, kriegte er dann eine schwere Ernährungsstörung.“

WhisperX: „Nach einer Woche oder so als Säugling, 14 Tage später, kriegte ich dann eine schwere Ernährungsstörung.“

[3] Schachtschneider, Olaf , Interview ev001, 03.06.2019, Band 1 – 0:00:42, Interview-Archiv "Eiserner Vorhang", https://archiv.eiserner-vorhang.de/de/interviews/ev001?tape=1&time=0h00m42s, 07.03.2024

Schwächen der ASR

4) Named Entities

OT: „Und wir sind zu einem Rechtsanwalt gegangen, dem Herrn de Maiziére. Bekannt.“

WhisperX: „Und wir sind zu einem Rechtsanwalt gegangen. Den Herrn, die mir sehr bekannt.”

[4] Schachtschneider, Olaf , Interview ev001, 03.06.2019, Band: 1 – 0:05:50, Interview-Archiv "Eiserner Vorhang", https://archiv.eiserner-vorhang.de/de/interviews/ev001?tape=1&time=0h05m50s, 16.01.2026.

Schwächen der ASR

5) Non- und paraverbale Kommunikation / Pausen

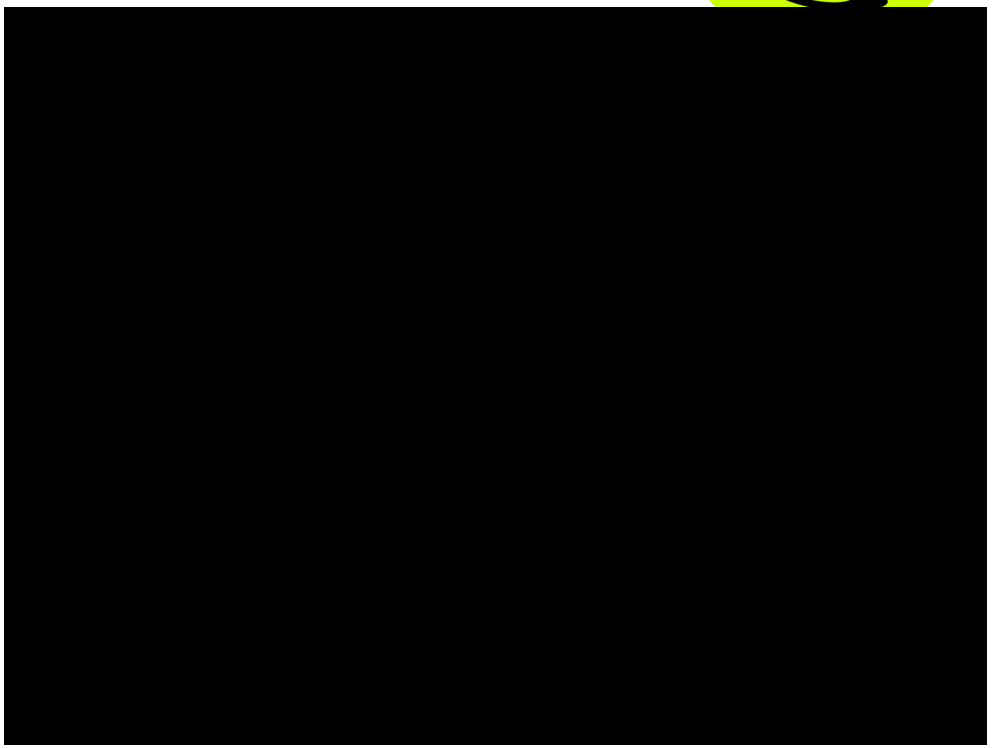
OT: „Ich sage: ‚Weißt du was, jetzt fahre ich mal schnell rüber, hole uns ein paar Zigaretten.‘ <s(lachend) Fahre die Schönhauser runter, bieg in die Brunnenstraße ein,> <g(darstellende Handbewegungen) da standen sie, einer nebeneinander.> <p4> Da bin ich nicht weitergekommen.“[5]

WhisperX: „Ich sage, weißt du was, jetzt fahre ich mal schnell rüber, hol uns ein paar Zigaretten. Ich fahre da schön runter und da kriege ich eine Brunnenstraße hin. Da standen sie jeder nebeneinander. Da bin ich nicht weitergekommen.“

[5] Schachtschneider, Olaf , Interview ev001, 03.06.2019, Band 1 – 0:11:23, Interview-Archiv "Eiserner Vorhang", https://archiv.eiserner-vorhang.de/de/interviews/ev001?tape=1&time=0h11m23s, 07.03.2024

Schwächen der ASR

6) Multilinguales Audio/Video



Simone G., Interview adg4560, null, Archiv „Deutsches Gedächtnis“, https://deutsches-gedaechtnis.fernuni-hagen.de/de/interviews/adg4560

Schwächen der ASR

7) Simultansprechen/ Sprecher*innenerkennung

IN	SPEAKER	TRANSCRIPT
00:00:00.082	SPEAKER_01	Ja, aber ist ja gut zu wissen, weil das ja auch quasi Interviews sind, die dann in die Sammlung irgendwie passen.
00:00:07.730	SPEAKER_00	Der Ton ist aus.
00:00:11.355	SPEAKER_00	Die Batterie ist leer von dem.
00:00:11.976	SPEAKER_01	Oh, das ist aber schlecht.
00:00:17.062	SPEAKER_01	Das können wir noch öffnen.
00:00:17.862	SPEAKER_01	Das machen wir nur mit dem Ton hier.
00:00:20.346	SPEAKER_01	Ja, genau.
00:00:20.966	SPEAKER_01	Wissen Sie, wir arbeiten ja... Reicht das?
00:00:25.111	SPEAKER_01	Hören wir Ihnen da gut?
00:00:25.591	SPEAKER_01	Das klingt ja noch nicht gut.
00:00:26.422	SPEAKER_01	Ich weiß nicht, ob Sie das schon gesehen haben, aber das Alumni-Büro hat Internetseiten, 70 Jahre FU Berlin, da werden immer vier Fragen gestellt und dann so ein kleiner Absatz,
00:00:40.428	SPEAKER_01	da gibt es Antworten dazu.
00:00:42.585	SPEAKER_01	Und die haben gesagt, Sie unterstützen uns, wenn wir uns verpflichten, Ihnen diese vier Fragen zu stellen.
00:00:47.652	SPEAKER_01	Dass Sie die gegebenenfalls aufschreiben können.
00:00:51.176	SPEAKER_01	Deswegen, das ist jetzt nicht von uns im Auftrag des Aluminiumberufs.
00:00:54.579	SPEAKER_00	Sondern jetzt kriege ich die, genau.
00:00:55.820	SPEAKER_01	Vielleicht so ganz kurz und knapp, also dass Sie das dann hätten.
00:01:02.246	SPEAKER_01	Was ist Ihnen aus Ihrer Zeit an der Freien Universität besonders in Erinnerung geblieben?
00:01:11.977	SPEAKER_00	Das ist, muss ich spuren, natürlich, wie verrückt.
00:01:16.721	SPEAKER_01	Hm.
00:01:19.359	SPEAKER_00	Naja, das ist doch der Malteserkreis.
00:01:21.325	SPEAKER_00	Und der Malteserkreis, dass wir auch tatsächlich dann die Bücher hingekriegt haben.

Lönnendonker, Dr. Siegward, Interview eg001, 20.03.2018, Band 1 – 4:53:19, Erlebte Geschichte - Freie Universität Berlin, https://archiv.erlebte-geschichte.fu-berlin.de/de/interviews/eg001?tape=1&time=4h53m19s, 07.07.2025

Schwächen der ASR

8) Halluzinationen

Häufiges Beispiel:

Stille im Audio → „Untertitel im Auftag des ZDF, 2017“

Lönnendonker, Dr. Siegward, Interview eg001, 20.03.2018, Band 1 – 4:53:19, Erlebte Geschichte - Freie Universität Berlin, https://archiv.erlebte-geschichte.fu-berlin.de/de/interviews/eg001?tape=1&time=4h53m19s, 07.07.2025

Schwächen der ASR

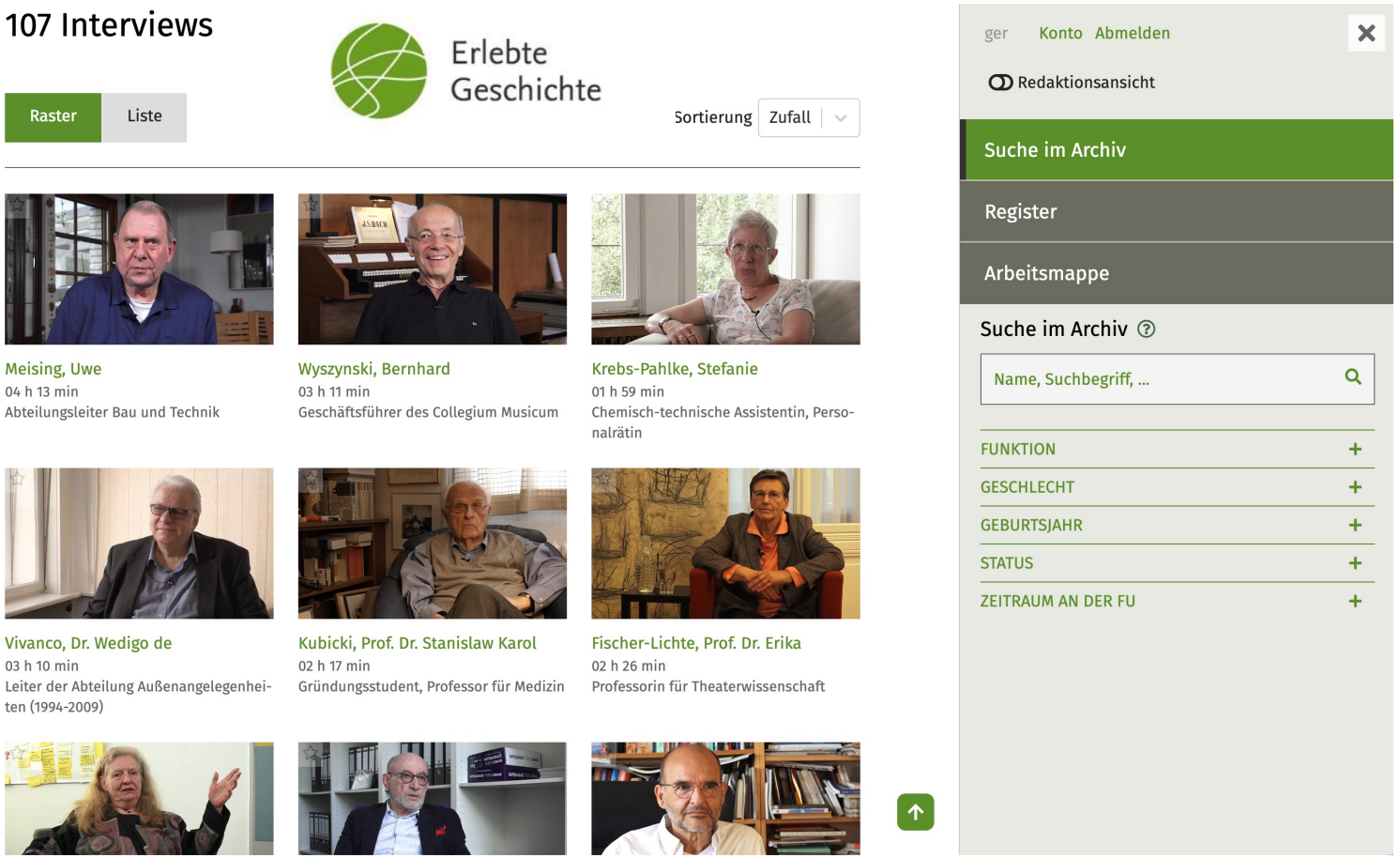
Zusammenfassung:

Whisper liefert keine wortgetreuen (oder gar lautgetreuen) Transkriptionen, sondern liefert Ergebnisse zwischen wortgetreu („verbatim“) und sinngemäß („gisted“).

Fine-tuning

Training mit Oral History-Interviews auf einem HPC Cluster

→ Ziel: Entwicklung eines domänenspezifischen Modells



Use case: 95 Interviews (>300h) zur Geschichte der Freien Universität Berlin mit standardisierten Transkripten

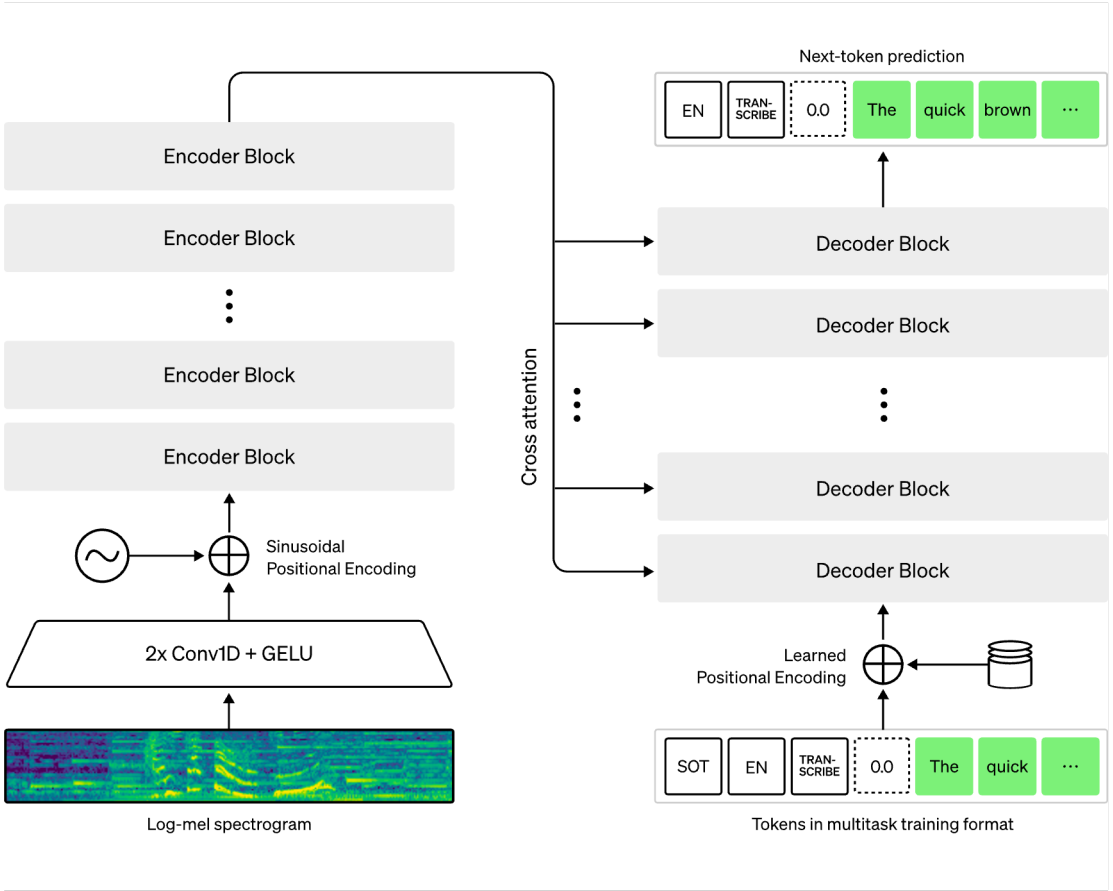
Whisper-Architektur

Sequence-to-sequence Transformer-Modell

Encoder: Übersetzung des Audio-Signals (als Log-mel spectogram) in eine maschinenlesbare Repräsentation

Decoder: „klassisches“ Sprachmodell, das den nächsten Token ermittelt, basierend sowohl auf der Audioinformation aus dem Encoder („cross-attention“) als auch des zuvor generierten Texts („self-attention“)

Large-v3: 1.55 B Parameter→ jeden Parameter finetunen?



Finetuning-Schritte:

Dataset preparation	Fine-tuning on HPC	Evaluation
Anonymisierung via GliNER und LLM Transformation in „Audiofolder”-Datensatzstruktur: 95 Interviews + Transkripte → 200.000 Audio-Snippets + metadata.csv Konvertierung in HDF5 format	Hyperparameter Fine-Tuning (Bayes’sche Optimierung, populationsbasiertes Training)	Quantitative Evaluation: WER-Berechnung Qualitative Evaluation: „LLM as a judge“

Demonstration:

https://www.fu-berlin.de/sites/ub/forschen/interviewsammlungen/forschung/ASR4Memory/Finetuning-Beispiele/index.html

WER-Vergleich

$$\text{WER} = \frac{\text{Zahl der ersetzten Wörter + Zahl der nicht transkribierten Wörter + Zahl der hinzugefügten Wörter}}{\text{Zahl der Wörter im Referenztranskript}}$$

Whisper Derivat	Modell	WER
Vanilla Whisper	large-v3	17,18%
Whisper Transformers	large-v3	17,8%
WhisperX	large-v3	17,7%
Whisper Timestamped	large-v3	17,1%
Whisper MLX	large-v3	18%
CrisperWhisper	large-v3 (fine-tuned)	21,2%
WhisperX	**large-v3 (fine-tuned)**	22.7%

„Unser“ Modell

Evaluation mit LLMs

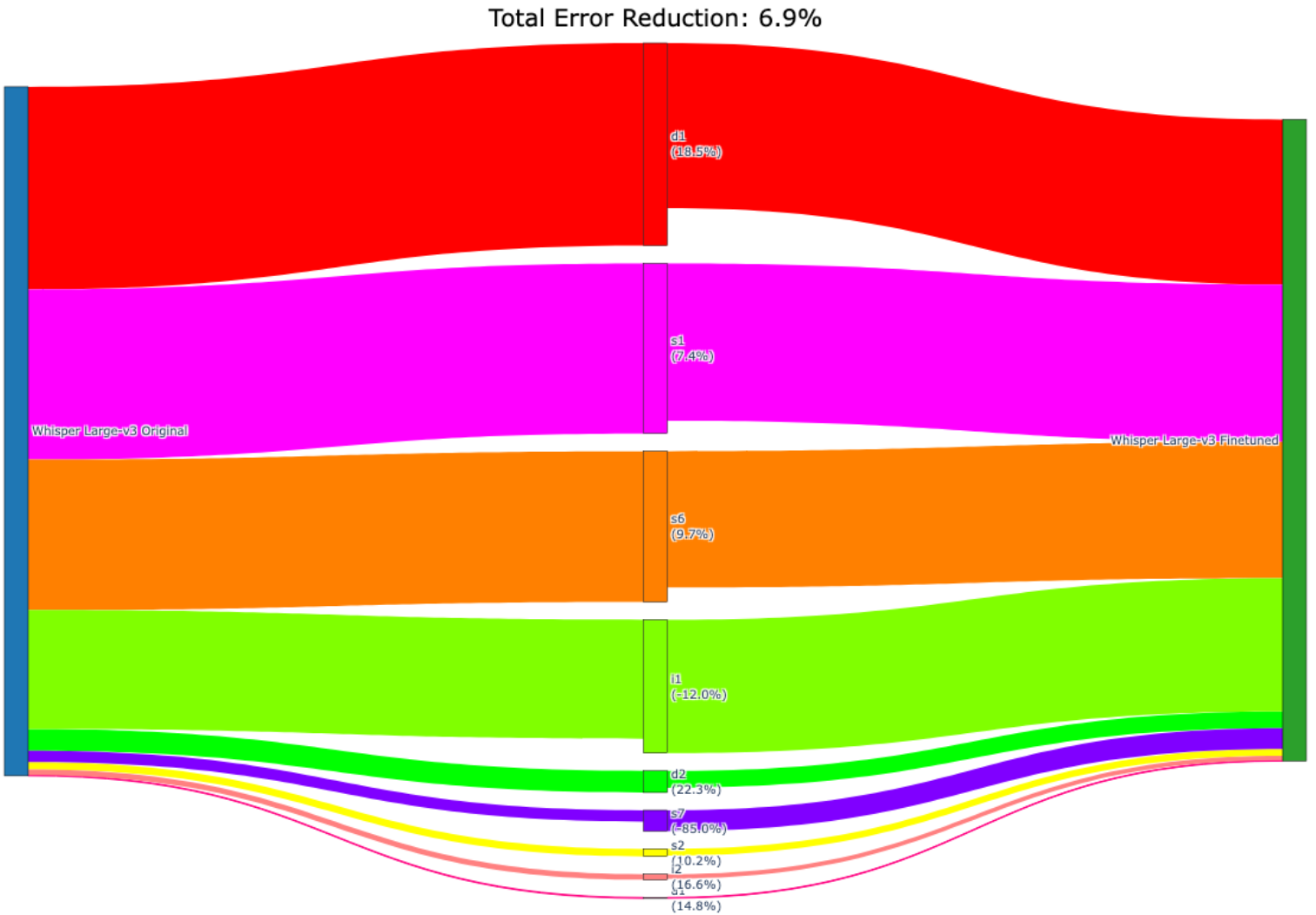
- WER als a quantative Maßeinheit enthält keine Informationen über Fehlertypen
- Qualitative Analyse ist sehr zeitaufwändig

→ automatische Analyse von Fehlertypen mithilfe von LLMs

Evaluation mit LLMs

	Fehlerkategorie	Fehlertyp	Definition	Relevanz	Beispiel	tag
0	Entfernen eines bedeutungslosen Wortes ohne in...	Deletion	Spracherkenner transkribiert ein inhaltlich be...	2	Artikel vor einem Nomen, z.B. „die" Frau oder ...	d1
1	Entfernen eines bedeutungsvollen Wortes mit in...	Deletion	Spracherkenner transkribiert ein inhaltlich be...	8	Personennamen, Orte, Ereignisse, für das Verst...	d2
2	Ersetzen eines bedeutungslosen Wortes ohne inh...	Substitution	Spracherkenner ersetzt im Transkript ein inhal...	2	Ersetzen des gesprochenen bedeutungslosen Wort...	s1
3	Ersetzen eines bedeutungsvollen Wortes mit inh...	Substitution	Spracherkenner ersetzt im Transkript ein inhal...	9	Ersetzen des gesprochenen bedeutungsvollen Wor...	s2
4	falsche oder fehlende Zuordnung eines Satzes z...	Substitution	Spracherkenner weist bei einem Sprecherwechsel...	2	Satz A wird von Sprecher A gesprochen, Satz B ...	a1
5	Veränderung des grammatischen Satzbaus	Substitution	Spracherkenner verändert im Transkript grammat...	5	Spracherkenner: "Ja also da war ich auch schon...	s6
6	Transformation eines mit Dialekt/Akzent gespro...	Substitution	Spracherkenner transformiert im Transkript die...	3	Spracherkenner: "Ich weiß, ich bin ein waschec...	s7
7	Einfügen eines halluzinierten bedeutungslosen ...	Insertion	Spracherkenner transkribiert ein nicht gesproc...	2	Spracherkenner: „ein"\nReferenztranskript: (St...	i1
8	Einfügen eines halluzinierten bedeutungsvollen...	Insertion	Spracherkenner transkribiert ein nicht gesproc...	10	Spracherkenner: „Untertitel von Arte Produktio...	i2

Error Flow and Reduction from Whisper Large-v3 Original to Finetuned



Zusammenfassung

Welche der zuvor aufgezählten Probleme konnten wir lösen?

Problem	Verbessert?
Füllwörter	Ja
Wiederholungen / Satzabbrüche / Verzögerungen	Ja
Dialekte	Ja (mehr als erwartet)
Named Entities	Teilweise (wahrscheinlich mehr Daten notwendig)
Non- und paraverbale Kommunikation / Pausen	Nein (kein Bestandtteil des Fine-Tuning-Prozesses)
Multilinguale Interviews	Nein (kein Bestandtteil des Fine-Tuning-Prozesses)
Simultansprechen/ Sprecher*innenerkennung	Nein (kein Bestandtteil des Fine-Tuning-Prozesses)
Wortgenauigkeit	Wurde schlechter

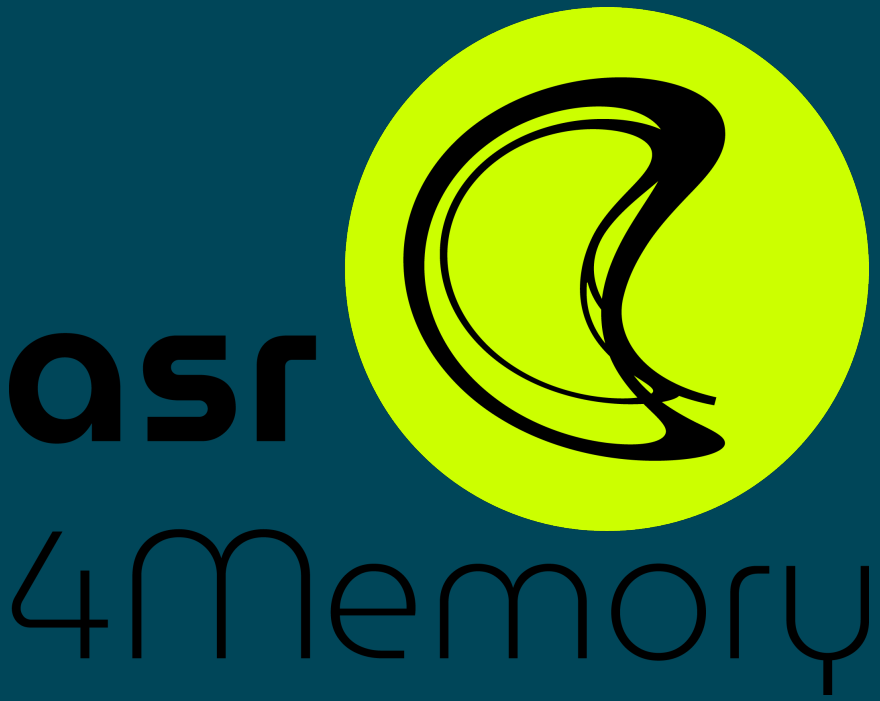
→ Wortgetreue behalten, neue Fehler minimieren

Repositorien

Datensatzvorbereitung: https://github.com/asr4memory/asr-dataset-creator

Finetuning und Evaluation: https://github.com/asr4memory/asr-finetune

Wie geht es weiter?



Nächste Schritte

Oral-History.Digital (oh.d)

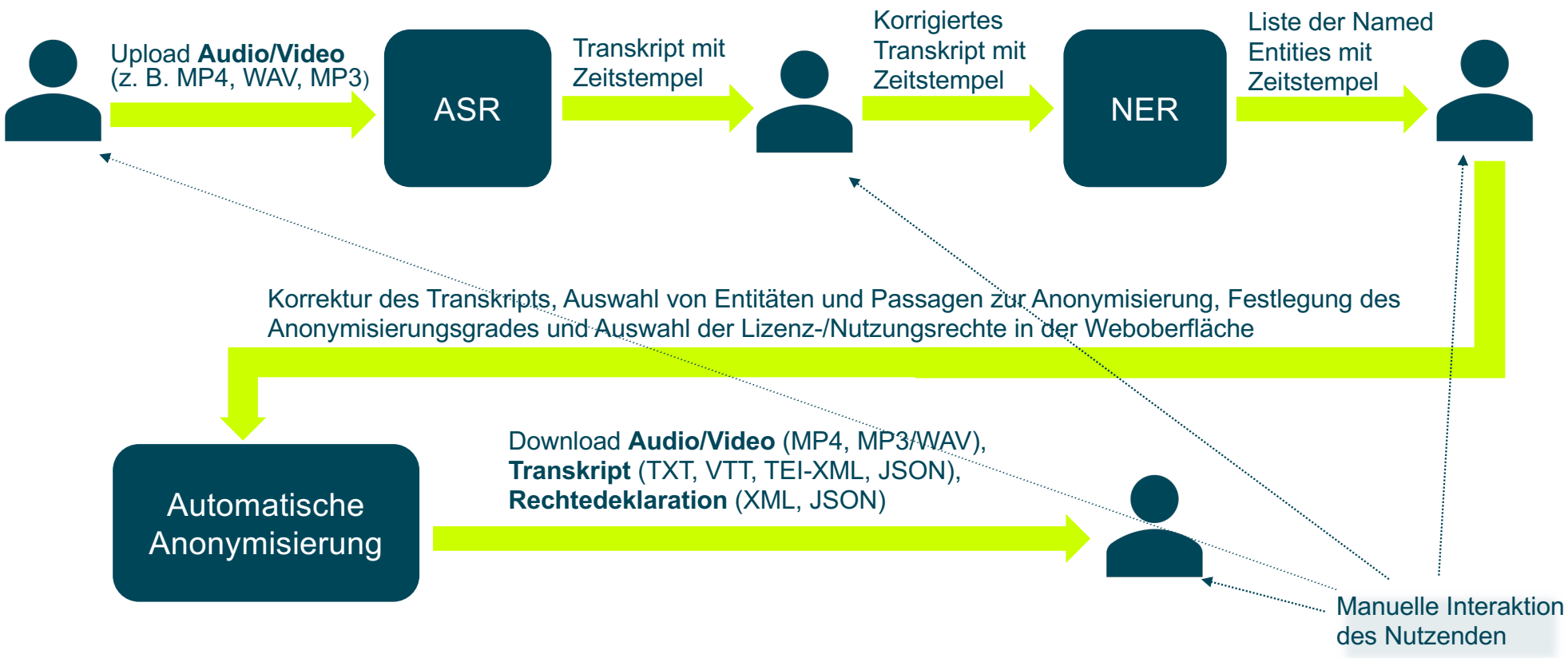
- **Schrittweise Integration der ASR-Funktionalität in die Erschließungs- & Recherche-Plattform**
- Start der ASR-Nutzung für oh.d-User seit 2025
  - Viele Anfragen, verschiedene Sprachen
  - Ziel: hohe Automatisierung der Workflows
  - https://portal.oral-history.digital/

Nächste Schritte

Open.Oral-History (o.oh)

- **Ausbau des Media Management Tools (MMT)**
- Ermöglichung von browserbasierten Korrekturen am Transkript
  - Entitätenerkennung (NER)
  - Anonymisierung der Transkripte und AV-Quellen
  - Ausgabe maschinenlesbarer Rechtedeklarationen
  - https://www.fu-berlin.de/ooh

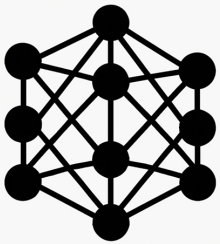
Projekt Open.Oral-History (O.OH)



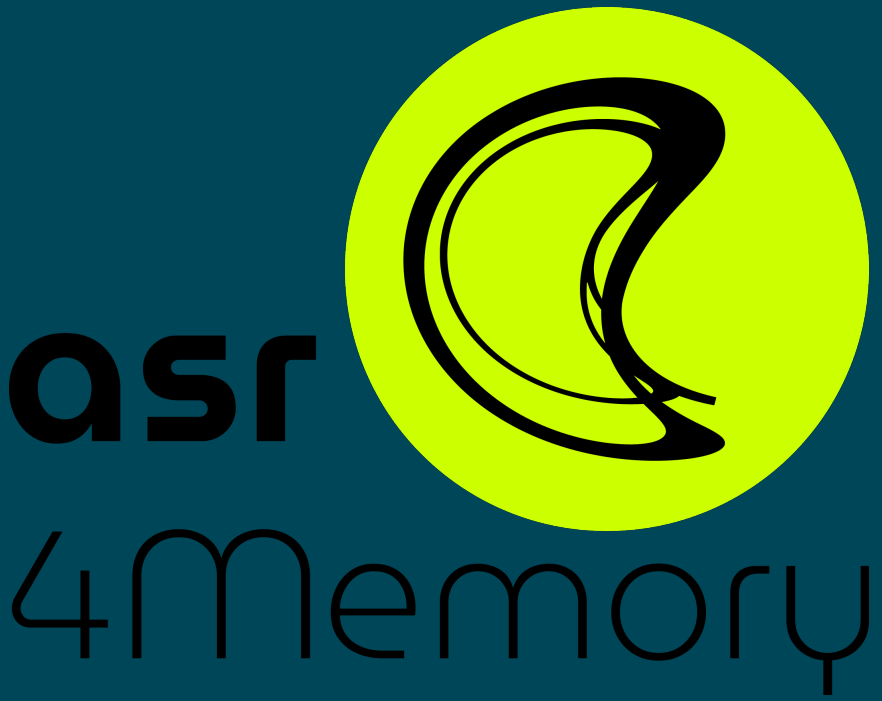
Nächste Schritte

ASR4Memory

- **Weiterentwicklung** der Transkriptions-Anwendung
  - Weitere KI-Komponenten in ASR-Pipeline einbinden, z. B. Zusammenfassung, Inhaltsverzeichnis, Metadatenextraktion
  - Zuverlässigkeit der Sprecherauszeichnungen verbessern
  - Mehrsprachigkeit (Code Switching) unterstützen (**Audio-Visual.Digital**)
  - ASR-Modell anpassen durch Fine-Tuning:
    - Glättungen und Halluzinationen reduzieren (**MATH+**, **KI.OH**)



Welche Anforderungen
und Wünsche haben Sie?



Rückmeldung

- **Wichtig ist Ihr Feedback!**
  - Welche Fehler oder Auffälligkeiten treten bei der ASR auf?
  - Welche Transkriptformate sind für die Nachnutzung wichtig?
  - Welche weiteren Funktionen werden gewünscht?
  - Welche Schnittstellen zu ASR4Memory wären interessant?
  - Etc.

Links

- **Projekt-Webseite:**
  https://www.fu-berlin.de/asr4memory

- **GitHub-Repositorien:**
  https://github.com/asr4memory

- **Media Management Tool (MMT)**
  https://mmt.oral-history.digital/

Kontakt

- asr@oral-history.digital
- tobias.kilgus@fu-berlin.de
- peter.kompiel@fu-berlin.de

Vielen Dank!