Introduction to Handwritten Text Recognition

Tobias Hodel (Walter Benjamin Kolleg, Digital Humanities)

Michigan, 13th May 2026

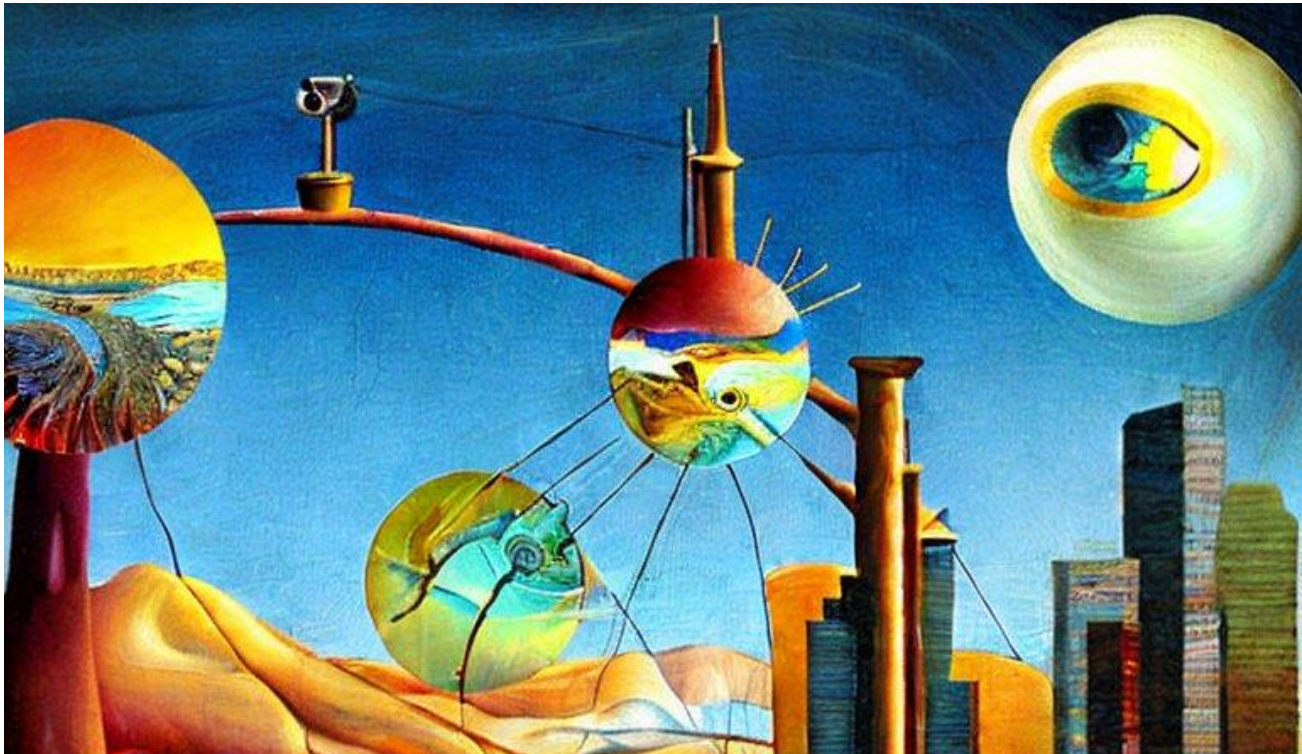


Bild: Stable Diffusion, Prompt:
«A deep neural network
in the style of a Dali painting»

Goals

We want to…

- get an Introduction to HTR/ATR
- Try out an Integrated Transcription Environment (ITE)
- Get to Know the Technology
- Compare it to Visual Language Models
- Discuss Best-Practices, Issues, etc.
- Develop our own approaches
- Get Access to Resources

Overview (Part I)

UNIVERSITÄT BERN

09:45	ATR: Where are we at the moment?
10:10	Round of Introductions: ATR — Where are you?
10:25	Key Concepts & Epistemological Decisions
10:40	Integrated Transcription Environments
10:55	How to use eScriptorium
11:20	eScriptorium vs. Transkribus
11:25	Business models, open science, and HTR-United

Overview (Part II)

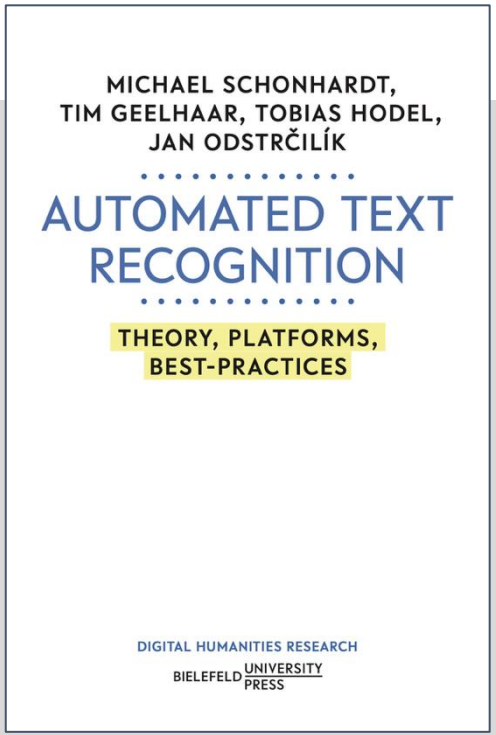
UNIVERSITÄT BERN

14:15	Test your own material
14:45	Show your results
15:05	Introduction to VLMs and command-line tools
15:25	Developing Workflows and Identifying Issues
15:45	Open Questions and Inputs on Demand

Preparation and Resources

You will need:

- Access to escriptorium.flow-project.net
    - User account: mail address
    - Password: Kzoo2026
- Documentation for today: https://thodel.github.io/atr/
- Slides: XXX



A very short overview: From Optical Character Recognition to (Automatic/Handwritten) Text Recognition

UNIVERSITÄT BERN

1990ies: Commercial Optical Character Recognition

2000thes: Recognition of «Fraktur» with OCR
(EU Project IMPACT, ABBYY XIX)

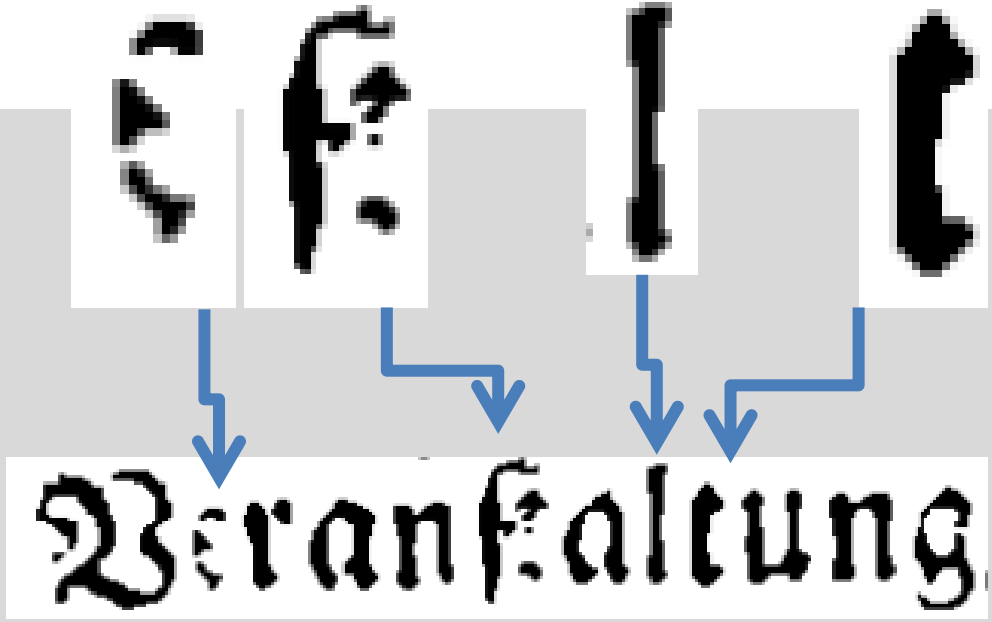
2010ns: Simple Language Model / Introduction of Neural Networks
(to the field)

After 2016: Dominance of *deep learning*
(high-dimensional language model)

2021: Transformer-based Models for text recognition (trOCR)

2024: Visual/Multimodal Language Models, based on LLMs

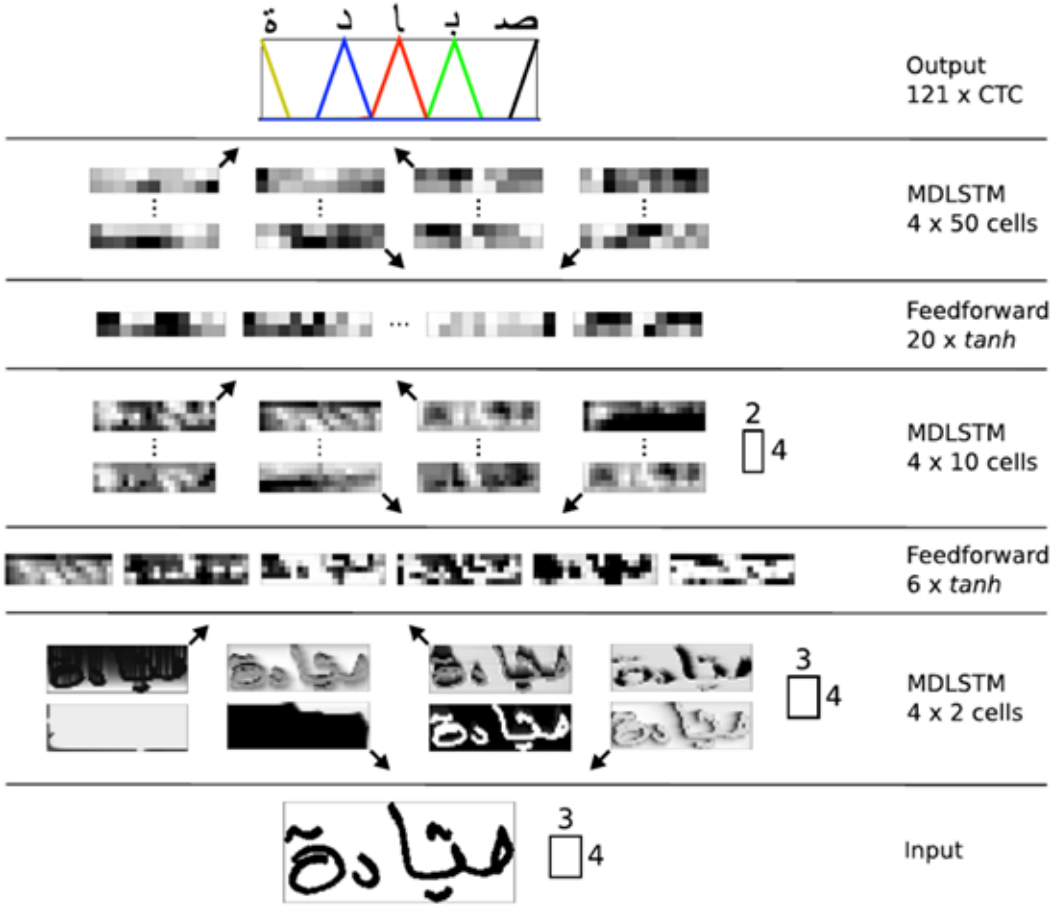
Automated Text Recognition?



Automated Text Recognition

UNIVERSITÄT BERN

- Machine learning using neural networks
- Processes writing by line, rather than by character
- Needs to be trained by being shown document images and transcripts
- More training data → more accurate recognition
- Create a model to transcribe and search a collection of documents



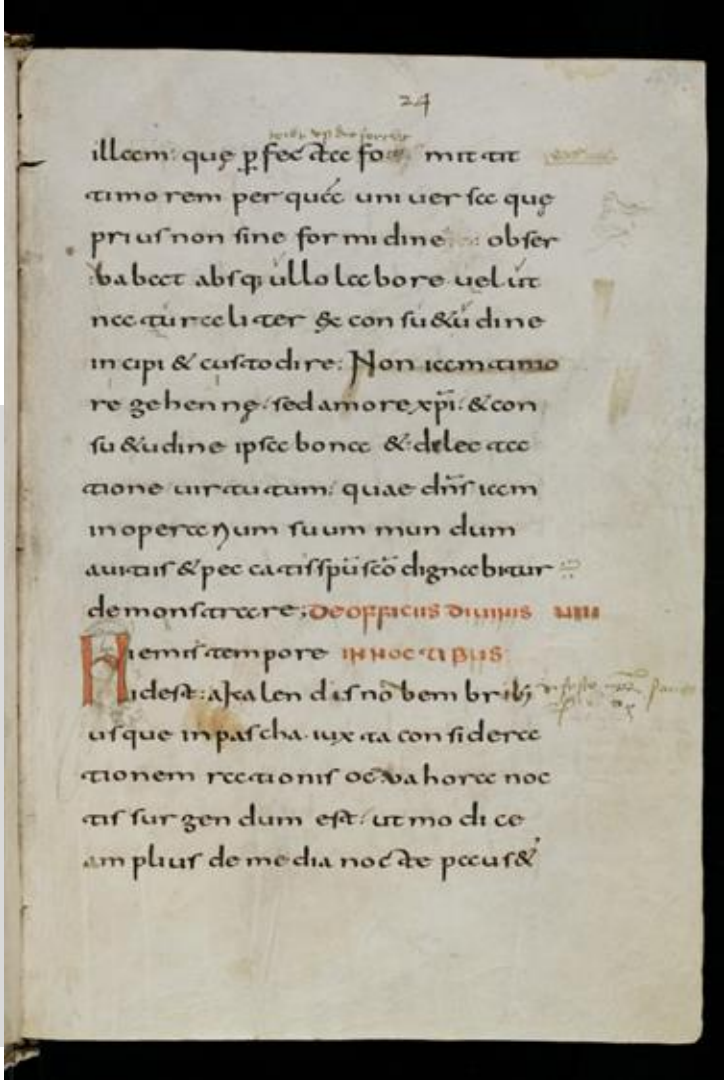
(Bild aus GRAVES, SCHMIDHUBER: Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks)

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.00	0.02	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.88	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.61	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Carolingian Minuscles

5'000 words, 1'000 lines:

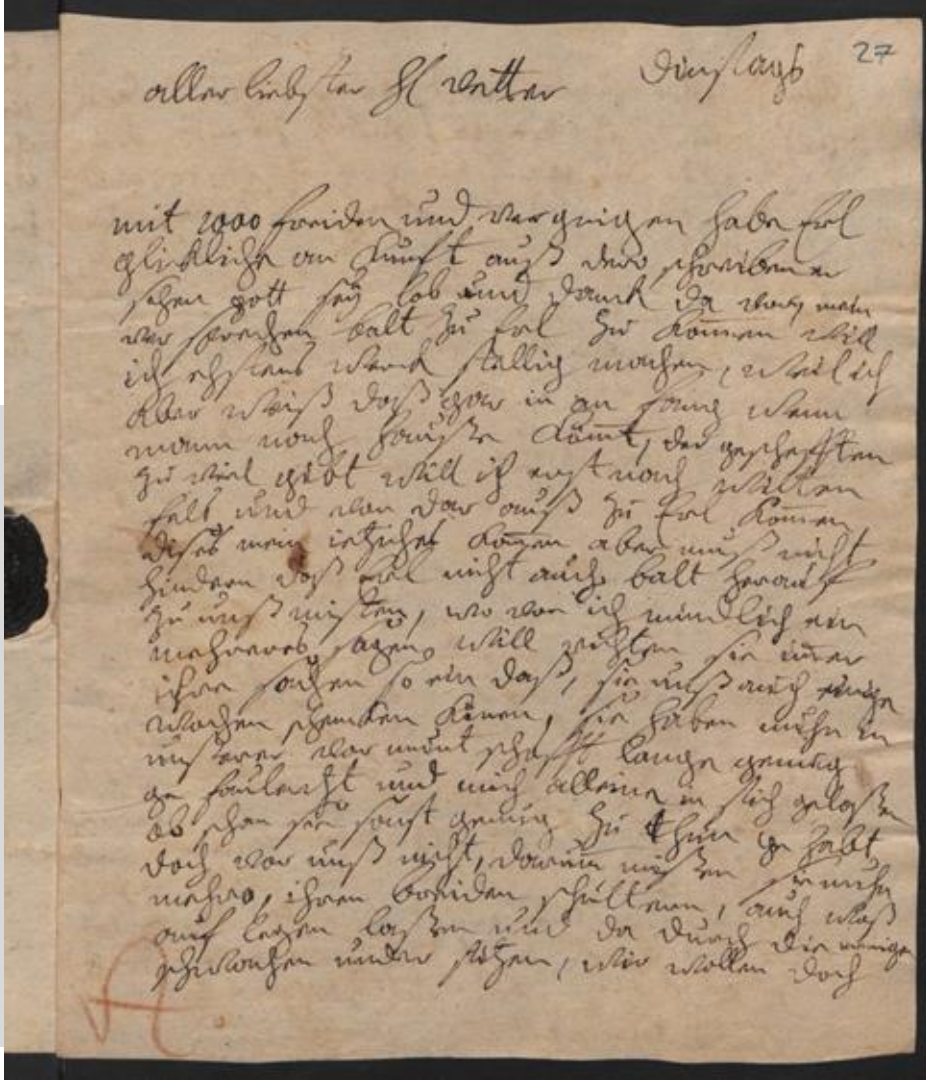
Validation pages: 7% CER



Early Modern Letters

48'000 words, 6'450 lines:

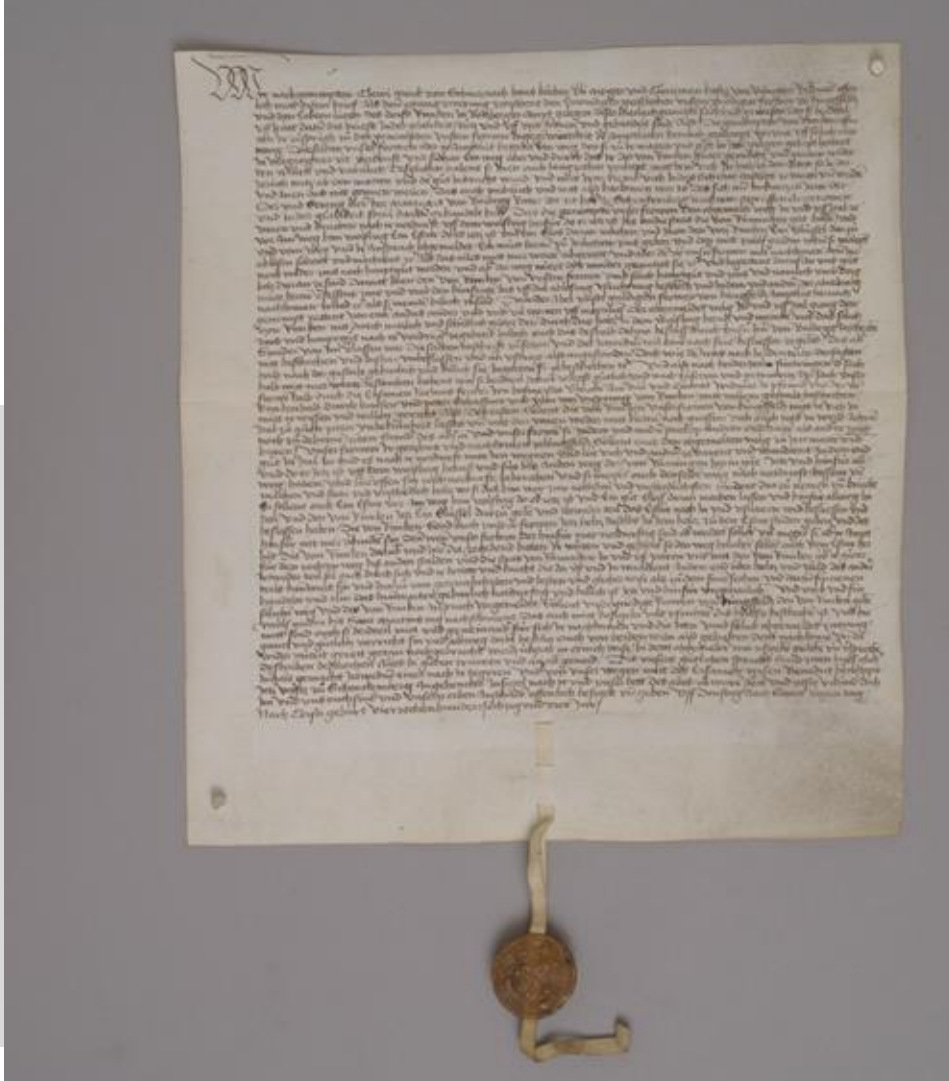
Validation pages: 2,5% CER



Charter
15th century

77'000 words, 3'500 lines:

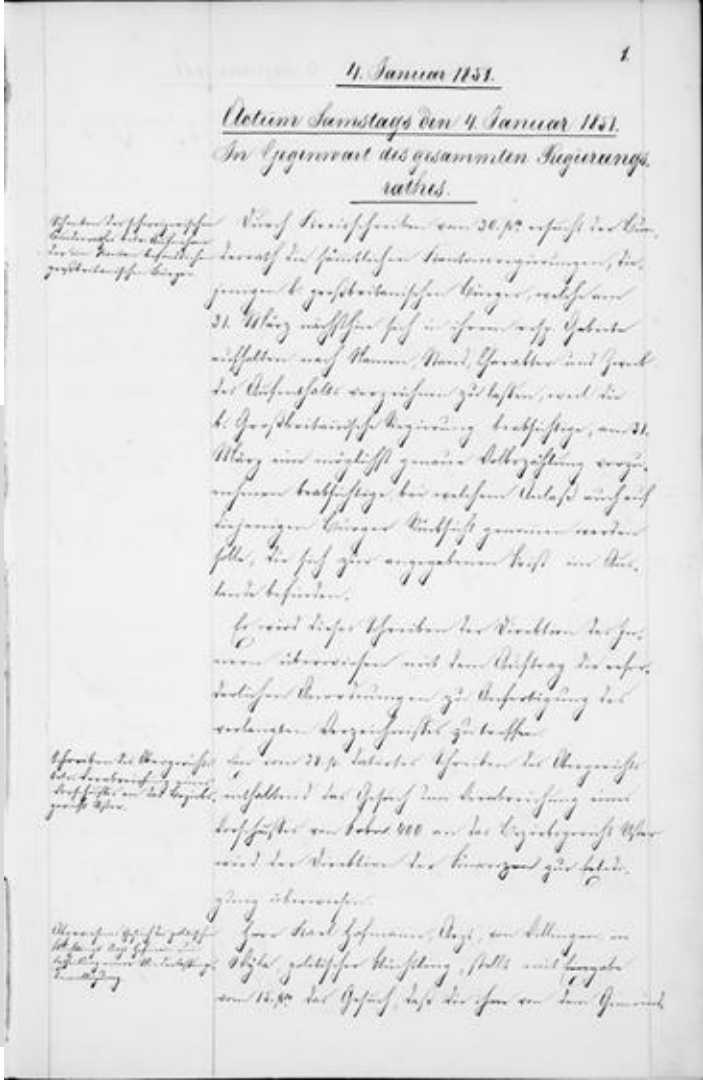
Validation pages: 4,8% CER



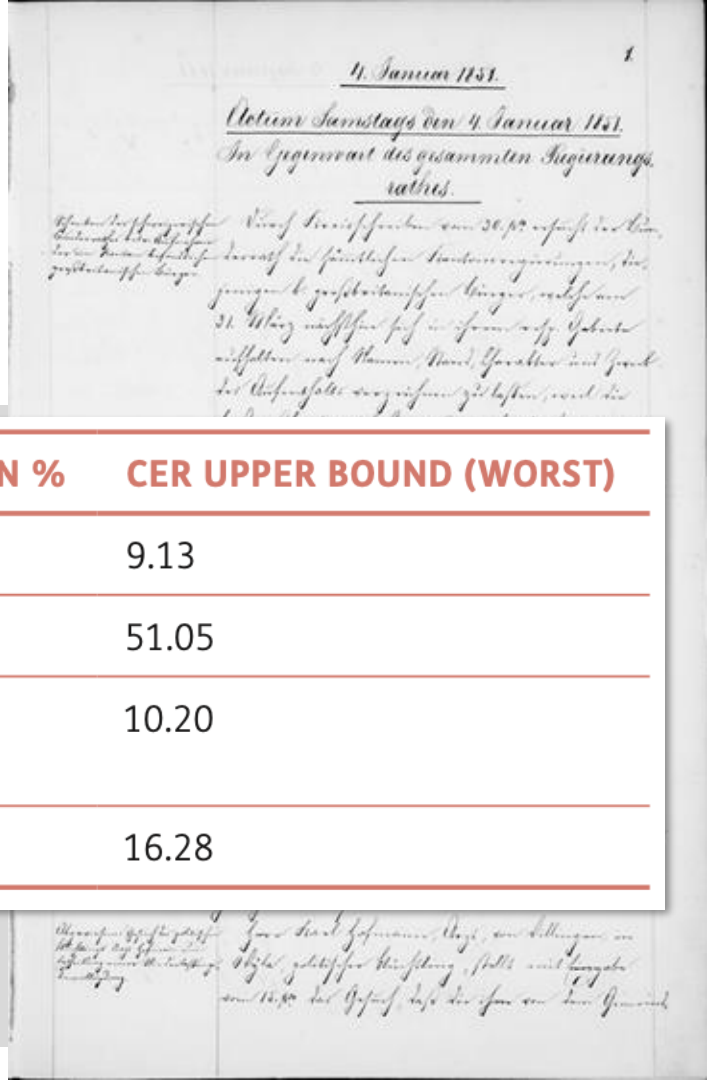
Council Minutes 19th Century

148'000 words, 31'000 lines:

2,5% CER



Generalizing of Writing Types Current/19th century

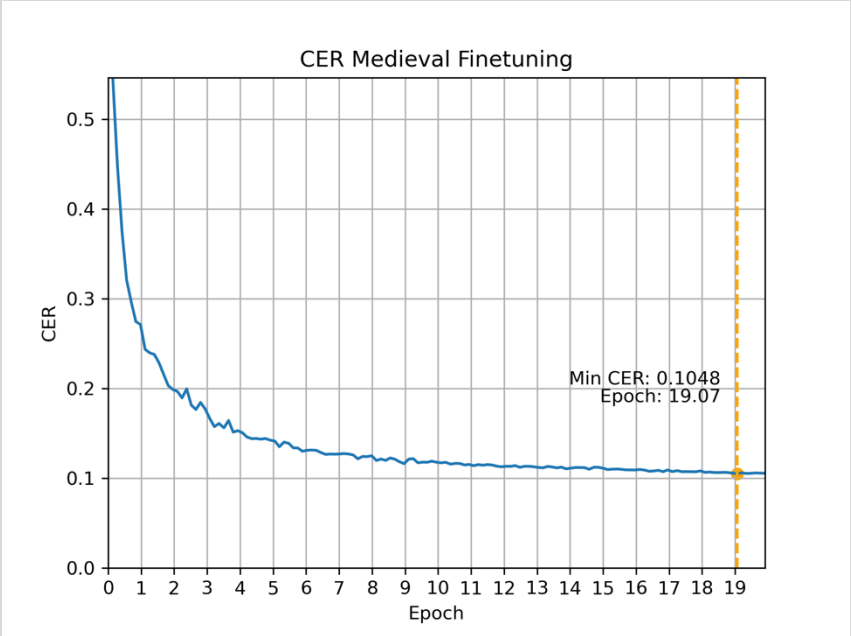


HTR MODEL	HTR ENGINE	CER MEAN %	CER MEDIAN %	CER UPPER BOUND (WORST)
German Kurrent M2	HTR+	3.43	2.76	9.13
	PyLaia	18.77	13.30	51.05
Transkribus German Kurrent	HTR+	5.90	4.85	10.20
RRB	HTR+	9.15	8.13	16.28

Generalizing of Writing Types

Medieval Scripts (TrOCR)

→ Transformer-based Models



What matters

Keep in mind…

- uniformity is key
- it's about your goals for a text/document
- there's not such a thing as a *perfect* transcription
- most machine learning algorithms start with random states; meaning: it doesn't hurt to re-run training processes

Transkribus, eScriptorium & Co.

Integrated Transcription Environments (ITE), Schonhardt et al. 2026, p. 37ff.

- Processing Platform (User, Login, Server Space)
- Sharing of Documents
- Creation of Training Material
- Run Training jobs (if infrastructure is available)

Visual Language Models (as an alternative)

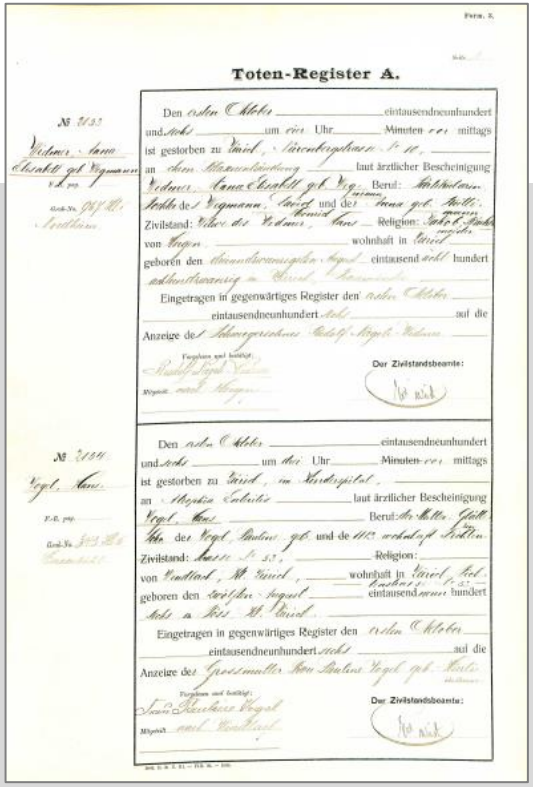
Based on latent space vectorization as in Large Language Models

Run recognition process on entire pages

- no segmentation of layout
- no training of specific handwritings
- relying on large closed- / open-providers (Mistral, QWEN, Google)

Test out with Gemini (most successful for historical handwriting):
https://gemini.google.com/app

VLM: Text



Toten-Register A.

Preprocessing (Segmenting in two entries) with YOLO v7 (specific model)

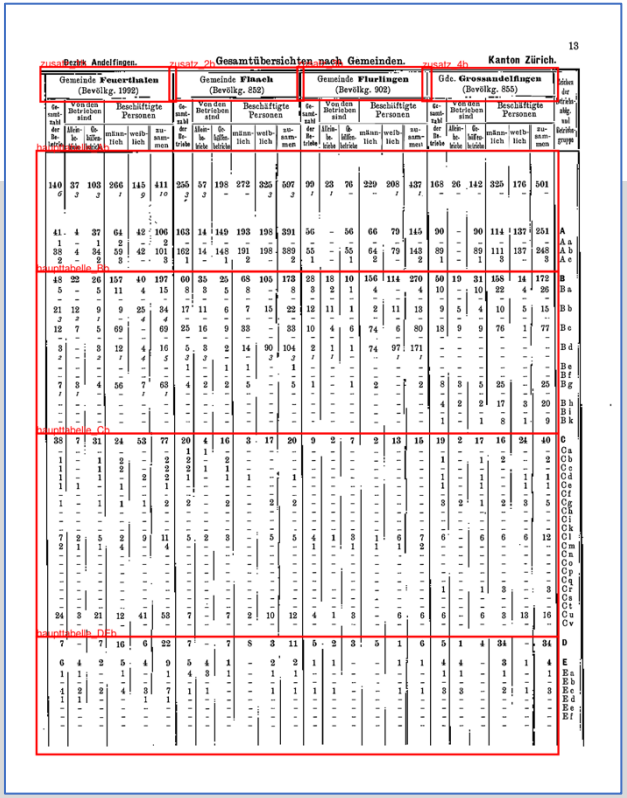
Text Recognition with QWEN 3 VL (8B)

→ for prints in GE after 1850 (handwritten and printed)

Information Extraction with GPT-OSS (120 B)

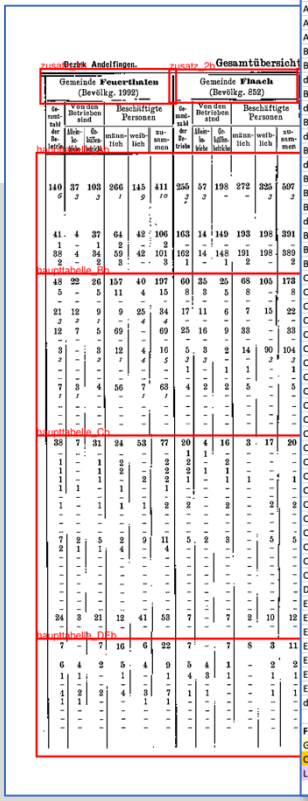
→ Segmenting in information pieces

VLM: Tables



- Identification of four/five segments
- Processing with MistralOCR
- Creation of a consolidated table with marked up sum errors

VLM: Table Recognition



Kategorie	A_col_1	A_col_2	A_col_3	A_col_4	A_col_5	A_col_6	A_col_7	A_col_8	A_col_9	A_col_10	A_col_11	A_col_12	B_col_1	B_col_2	B_col_3	B_col_4	B_col_5	B_col_6	B_col_7	B_col_8	B_col_9	B_col_10	B_col_11	B_col_12	B_col_13	B_col_14	B_col_15	B_col_16	B_col_17	B_col_18
Gesamtsum	154	52	102	203	183	386	457	140	317	993	929																			
darunter hi	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A. Gewinnu	93	9	84	166	94	260	184	14	170	389	171																			
Aa. Bergbau	2	-	2	1	-	1	1	-	1	-	-																			
Ab. Landwi	91	9	82	165	94	259	180	14	166	282	171																			
Ac. Forstwi	-	-	-	-	-	-	3	-	3	107	-																			
B. Veredlur	44	36	8	25	74	99	174	99	75	435	627		76	52	24	51	68	99	133	96	37	152	243	395	93	63	30	65	135	200
B a. Herstel	2	-	2	3	-	3	22	2	20	70	31		2	-	2	2	3	5	6	-	6	11	4	15	7	-	7	7	1	8
darunter hi	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B b. Herstel	7	6	1	4	3	7	51	34	17	30	47		13	9	4	6	9	15	25	20	5	12	14	26	22	14	8	5	17	22
darunter hi	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bc. Herstel	6	3	3	4	1	5	32	9	23	150	2		8	3	5	15	-	15	30	15	15	43	1	44	15	5	10	21	3	24
darunter hi	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B d. Herstel	27	25	2	12	70	82	52	50	2	136	680		48	39	9	1	56	57	62	55	7	71	223	294	42	41	1	28	114	142
darunter hi	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B e. Herstel	-	-	-	-	-	-	1	-	1	7	1		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B f. Chem. H	-	-	-	-	-	-	-	-	-	-	8		-	-	-	-	-	-	-	-	-	-	-	3	1	2	-	-	-	-
B g. Bearbe	2	-	2	-	2	11	3	8	18	-			4	1	3	5	-	5	9	5	4	15	-	15	4	2	2	4	-	4
darunter hi	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B h. Verviel	-	-	-	-	-	-	4	1	3	18	2		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B i. Wassern	-	-	-	-	-	-	1	-	1	6	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B k. Gewerk	-	-	-	-	-	-	-	-	-	-	6		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C. Handel	14	4	10	10	14	24	76	15	61	69	90		11	1	10	9	12	21	36	10	26	30	44	74	25	6	19	10	5	19
Ca. Lebend	-	-	-	-	-	-	6	6	-	-	-		1	1	-	1	-	1	4	4	-	-	-	-	2	1	1	1	-	1
C b. Rohprc	-	-	-	-	-	-	2	-	2	4	1		-	-	-	-	-	-	2	-	2	-	-	-	2	2	-	1	-	1
C c. Landwi	3	1	2	6	-	6	1	-	1	1	1		1	-	1	1	-	1	1	-	1	-	-	-	3	-	3	-	-	-
C d. Stein-,	-	-	-	-	-	-	4	1	3	4	3		1	-	1	1	1	2	-	-	-	-	-	-	1	-	1	2	1	3
C e. Holzwa	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	1	-	1	1	1	2	-	-	-	-	-	-
C f. Buch-, K	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C g. Spinnst	-	-	-	-	-	-	7	-	7	8			1	-	1	-	1	1	5	1	4	6	7	13	-	-	-	-	-	-
C h. Ausstei	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C i. Warenh	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C k. Bazars	-	-	-	-	-	-	2	-	2	4	3		3	-	3	3	2	5	9	4	5	2	8	10	9	1	8	2	4	6
C l. Lebensi	4	3	1	-	2	2	21	6	15	13	19		-	-	-	-	-	-	2	-	2	-	1	1	1	1	-	-	1	1
C m. Geträr	-	-	-	-	-	-	6	1	5	15	2		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C n. Tabak u	-	-	-	-	-	-	2	1	1	3	2		-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	1	-	1
C o. Drogue	-	-	-	-	-	-	2	-	2	2	2		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C p. Abfälle	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C q. Trödler	-	-	-	-	-	-	-	-	-	-	-																			
C r. Bankwe	-	-	-	-	-	-	1	-	1	3	3																			
C s. Versich	-	-	-	-	-	-	1	-	1	2	2																			
C t. Vermitt	-	-	-	-	-	-	-	-	-	-	-																			
C u. Gastwi	7	-	7	4	12	16	21	-	21	11	52																			
C v. Autom	-	-	-	-	-	-	-	-	-	-	-																			
D. Verkehr	triebe	100 triebe	1000-betrie	lich	lich	Stimmen	100 triebe	1000-betrie	lich	lich	Stimmen		4	-	4	26	6	32	7	-	7	13	5	18	5	-	3	8	6	16
E. Öffentl. V	1	2	100	200	100	200	100	200	100	200	1000		1	1	-	-	1	1	6	4	2	5	2	7	1	1	-	-	1	1
E a. Öffentl	2	2	-	2	-	2	6	2	6	77	7		-	-	-	-	-	-	2	-	2	3	-	3	-	-	-	-	-	-
E b. Rechtsi	1	1	-	-	1	1	15	10	5	23	31																			
E c. Gesund	-	-	-	-	-	-	1	-	1	-	-		1	1	-	-	1	1	4	4	-	2	2	4	1	1	-	-	1	1
E d. Unterri	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E e. Übrige	1	1	-	-	1	1	10	6	4	20	33		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E f. Künste	-	-	-	-	-	-	2	2	-	1	1																			
darunter hi	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

FARBLEGENDE:

Gelb	Zelle war leer/Null (0 eingefügt)
Orange	Vertikaler Summenfehler (Spaltensumme passt nicht)
Lila	Horizontaler Summenfehler (Zeilensumme passt nicht)

Automatic Text Recognition: Strategy

- Small Amount of Data without Model (pre-1800)

→ Transkribus

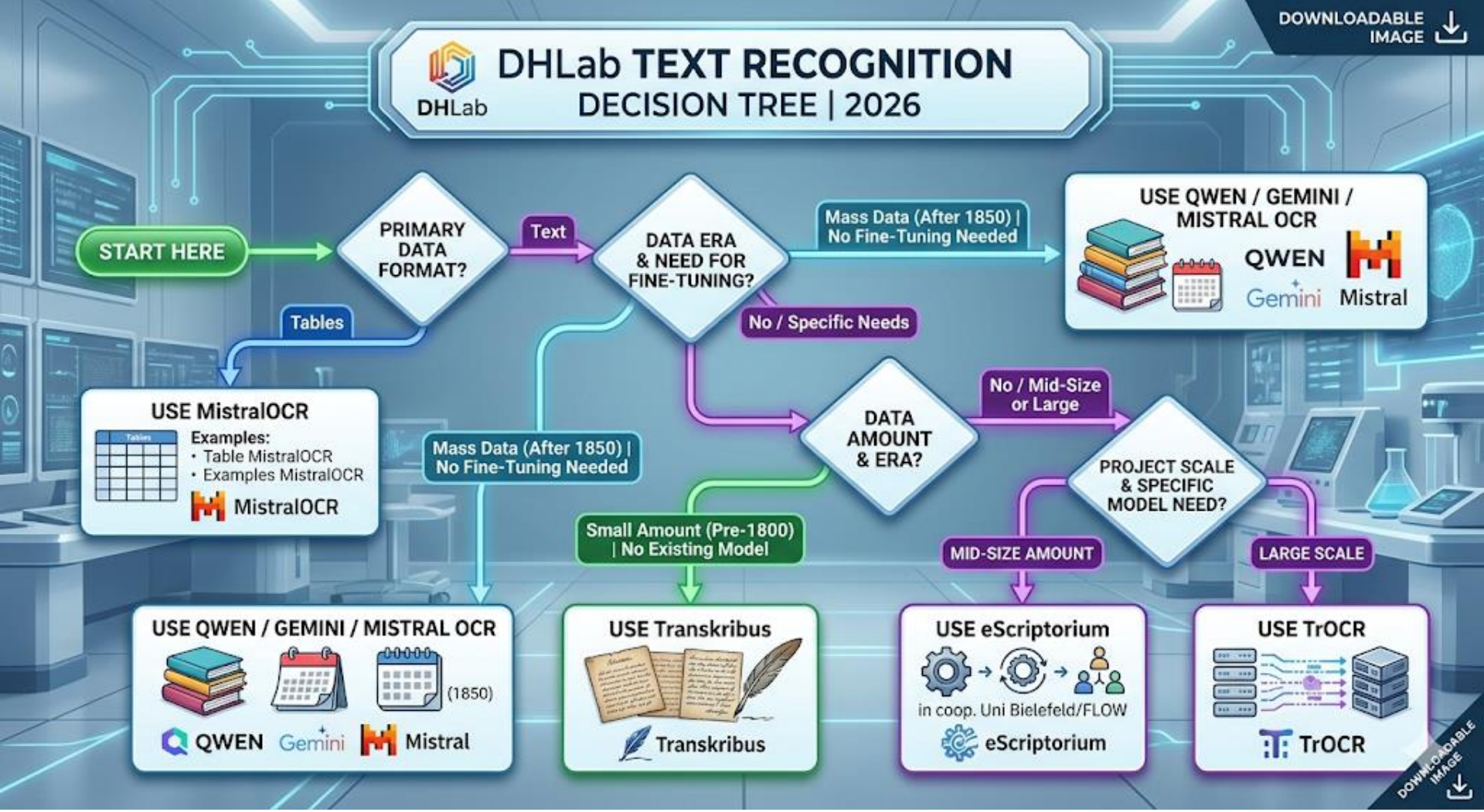
- Mid-Size Amount of Data with need for specific model

→ eScriptorium (in coop. Uni Bielefeld/FLOW)

- Large Scale projects with specific model

→ TrOCR

- Mass data after 1850, without need for fine-tuning → QWEN/Gemini/Mistral OCR
- Tables → MistralOCR


DHLab TEXT RECOGNITION
DHLab
DECISION TREE | 2026
DOWNLOADABLE IMAGE
START HERE
PRIMARY DATA FORMAT?
Text
Tables
DATA ERA & NEED FOR FINE-TUNING?
Mass Data (After 1850) | No Fine-Tuning Needed
No / Specific Needs
USE QWEN / GEMINI / MISTRAL OCR
QWEN
Gemini
Mistral
USE MistralOCR
Examples:
• Table MistralOCR
• Examples MistralOCR
MistralOCR
Mass Data (After 1850) | No Fine-Tuning Needed
DATA AMOUNT & ERA?
No / Mid-Size or Large
Small Amount (Pre-1800) | No Existing Model
PROJECT SCALE & SPECIFIC MODEL NEED?
MID-SIZE AMOUNT
LARGE SCALE
USE QWEN / GEMINI / MISTRAL OCR
(1850)
QWEN
Gemini
Mistral
USE Transkribus
Transkribus
USE eScriptorium
in coop. Uni Bielefeld/FLOW
eScriptorium
USE TrOCR
TrOCR
DOWNLOADABLE IMAGE

Round of Introduction

Where are you in regard to HTR

- Your name, background, pronouns
- Material you are working with
- Goals for this day
- Long-term application
- Experience so far

Key Concepts

… and why they matter

- Machine Learning
- Supervised Deep Learning
- Ground Truth
- Character Error Rate
- Word Error Rate
- Accuracy
- Character Set

Integrated Transcription Environments

Preparing, Training, Recognizing

- Graphical User Interface
- Set-up Once, Runs for Many
- Cooperate

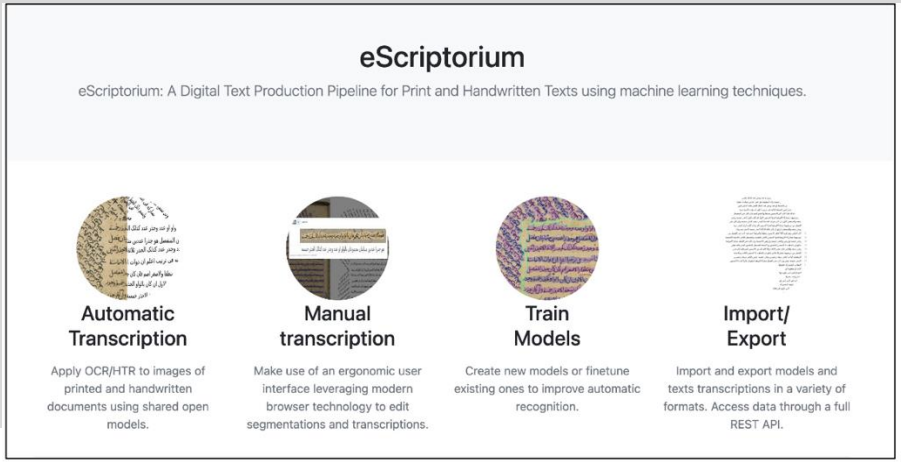
Not necessarily

- Open Source
- Reusability of Models

eScriptorium

An FOSS ITE

- Graphical User Interface
- Set-up Once, Runs for Many
- Cooperate
- Open Source
- Reusability of Models



eScriptorium

What is eScriptorium

- **Web-based tool** for transcription (no local installation necessary)
- **Free and Open Source** (FOSS, Code available on GitLab)
- **Developed by:** Université Paris Sciences et Lettres (PSL) as part of the projects *Scripta* and *RESILIENCE*.
- **Alternative to Transkribus:** Offers different features and is non-commercial.

eScriptorium

What is eScriptorium

- **Digital Text Production Pipeline:** Uses machine learning for print and handwritten texts.
- **Automatic Transcription:** Apply OCR/HTR to images using shared open models.
Manual Transcription: Ergonomic user interface to edit segmentations and transcriptions.
- **Train Models:** Create new models or fine-tune existing ones to improve recognition.
- **Import/Export:** Support for various formats and access through a full REST API.

eScriptorium

What is eScriptorium

Compatible with (can be imported):

- Page-XML
- Alto-XML
- IIIF-Manifest
- PDF
- Images

```
<?xml version="1.0" encoding="UTF-8"  standalone="yes"?>
<PcGts xmlns="http://schema.primaresearch.org/PAGE/gts/pagecontent/2019-07-15" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLo
  <Metadata>
        <Creator>escriptorium</Creator>
        <Created>2024-01-22T15:21:00.961404+01:00</Created>
        <LastChange>2024-01-24T12:23:49.148807+01:00</LastChange>
  </Metadata>
  <Page imageFilename="typewriter.jpeg" imageWidth="1400" imageHeight="1980">
    <TextRegion id="eSc_textblock_da7af2e3"  custom="structure {type:Main;}">
      <Coords points="47,300 47,652 1231,652 1229,307"/>
      <TextLine id="eSc_line_59a6603c" custom="structure {type:default;}">
        <Coords points="92,204 92,216 1144,222 1145,204 1145,179 92,178 92,204"/>
        <Baseline points="94,206 1147,206"/>
        <TextEquiv>
          <Unicode>Why Writing On A Typewriter Feels So Strangely Modern</Unicode>
        </TextEquiv>
      </TextLine>
      <TextLine id="eSc_line_7d9fec4c" custom="structure {type:default;}">
        <Coords points="92,237 92,250 236,260 255,252 279,260 374,252 399,262 562,262 574,252 648,260 733,252 744,260 787,252 815,260 846,252 870,
        <Baseline points="93,239 1107,239"/>
        <TextEquiv>
          <Unicode>Years of voice dictation trained me to think before I write</Unicode>
        </TextEquiv>
      </TextLine>
```

eScriptorium

is a tool for

- **Optical Character Recognition (OCR)** and **Handwritten Text Recognition (HTR)**
- **Automatic segmentation** and transcription
- **Manual transcription:** Used, for example, to produce **Ground Truth** (material for training models)
- **Training and Fine-tuning:** Powered by the free-and-open-source **Kraken** OCR Python engine in the background
- **Data Management:** Import, export, and cooperation features

eScriptorium

Demo

The eScriptorium instance can be reached at:
https://escriptorium.flow-project.net

eScriptorium

Demo

The eScriptorium instance can be reached at:
https://escriptorium.flow-project.net

eScriptorium

Strengths (+)	Weaknesses (-)
Free and Open Source	Software development can be slow and buggy
Opportunity for further development	Web-based tools can provoke restrictions (e.g., browser compatibility)
Great line mask recognition thanks to Kraken	No tagging/abbreviation features (requires external tools like Inception)
Can be self-hosted/operated	Few segmentation models available; prone to errors with complex layouts
Community models available (via Zenodo.org)	

Training

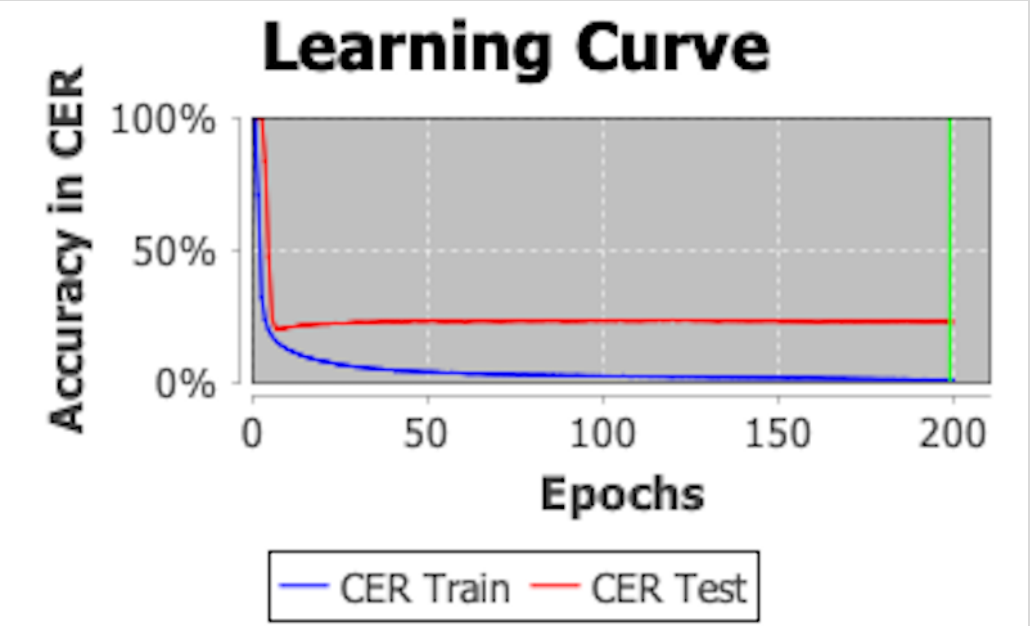
Create your own model

Necessities:
- Training Material
- Infrastructure
- Idea of ML approaches

Experience
- Interpretation of Training Curves

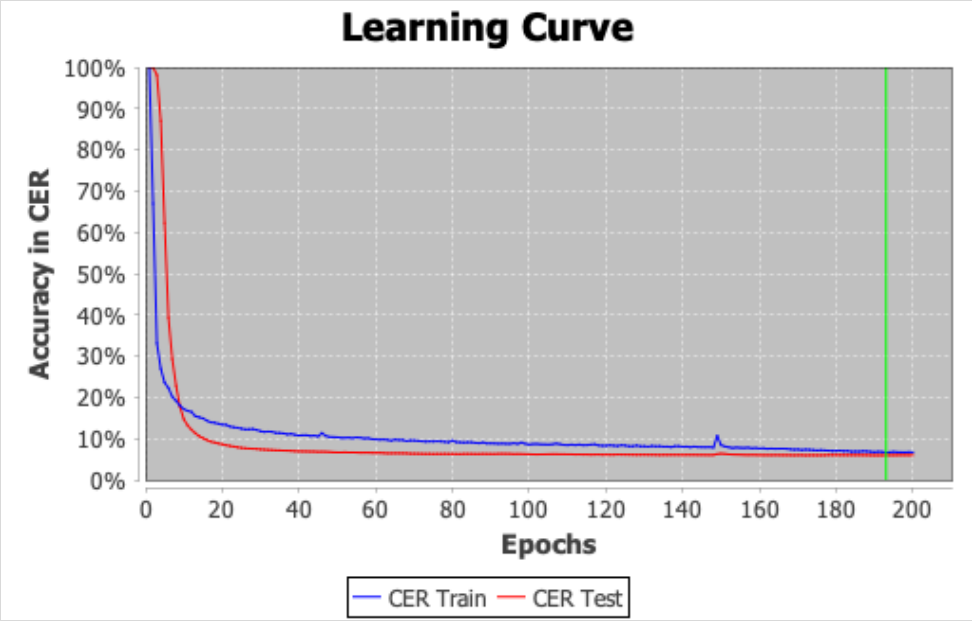
Interpretation of Graphs

Diverging curves: overfitting



Interpretation of Graphs

Fully trained/optimal network

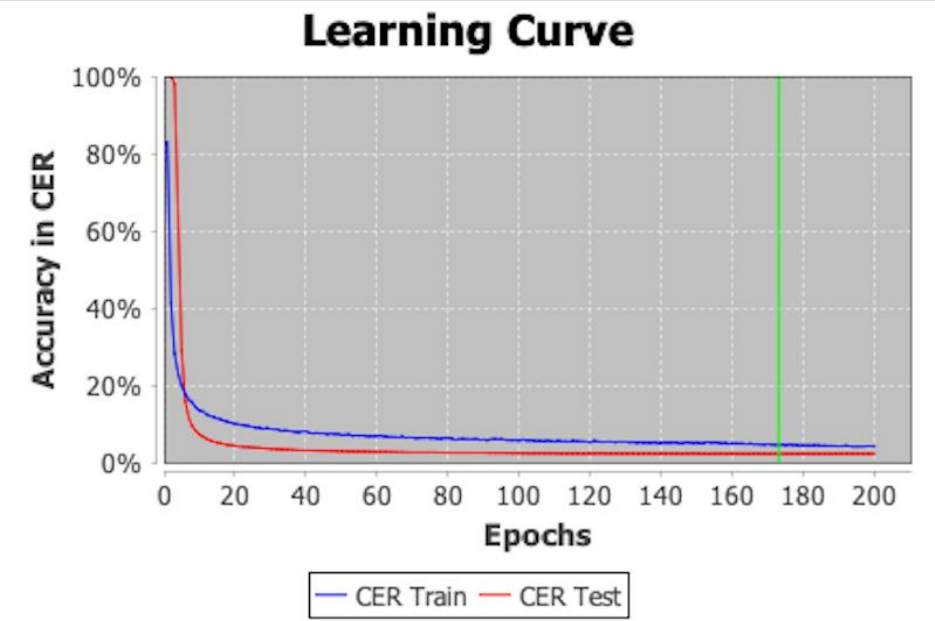


§

UNIVERSITÄT BERN

Interpretation of Graphs

Keep on training…



How to achieve good results

- define goals beforehand…
- … adapt your stile of transcription accordingly

  → Adherence to community guidelines

How to achieve good results

Transcription Guidelines:

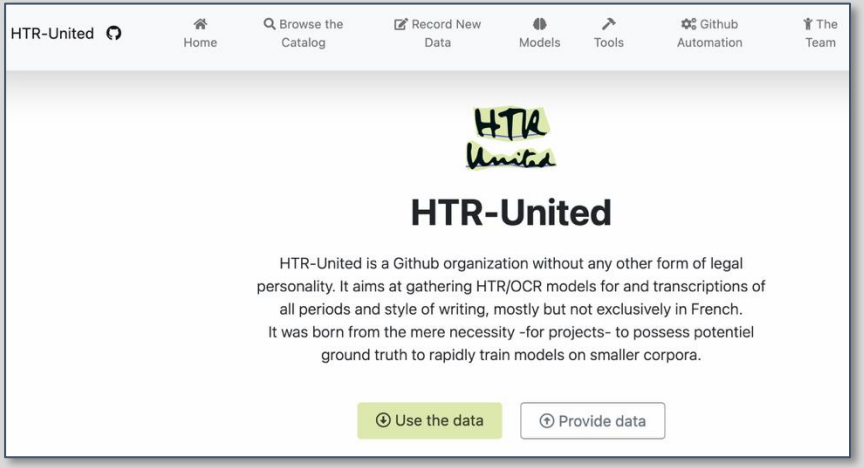
- CATMUS: https://hal.science/hal-03697382

Segmentation Ontology:

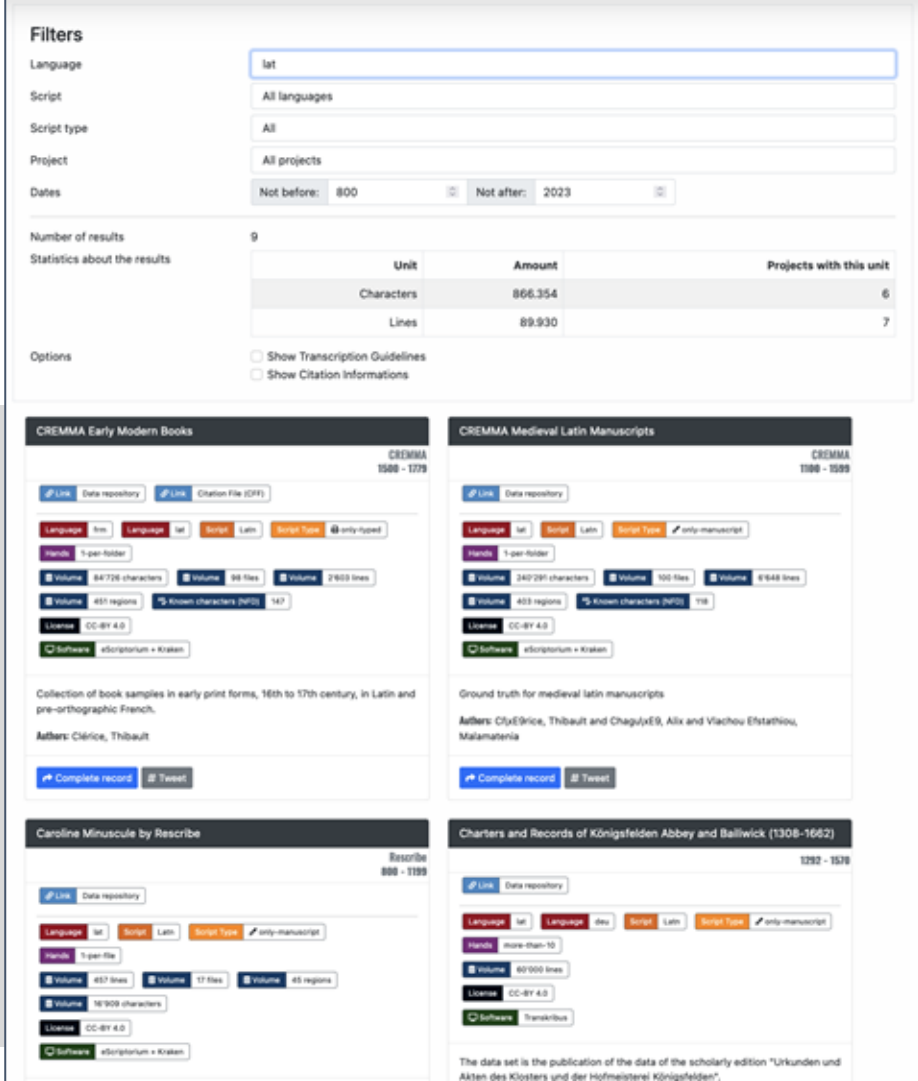
- https://jdmdh.episciences.org/14953
- https://segmonto.github.io/

Cooperation

Data Sharing: HTRunited

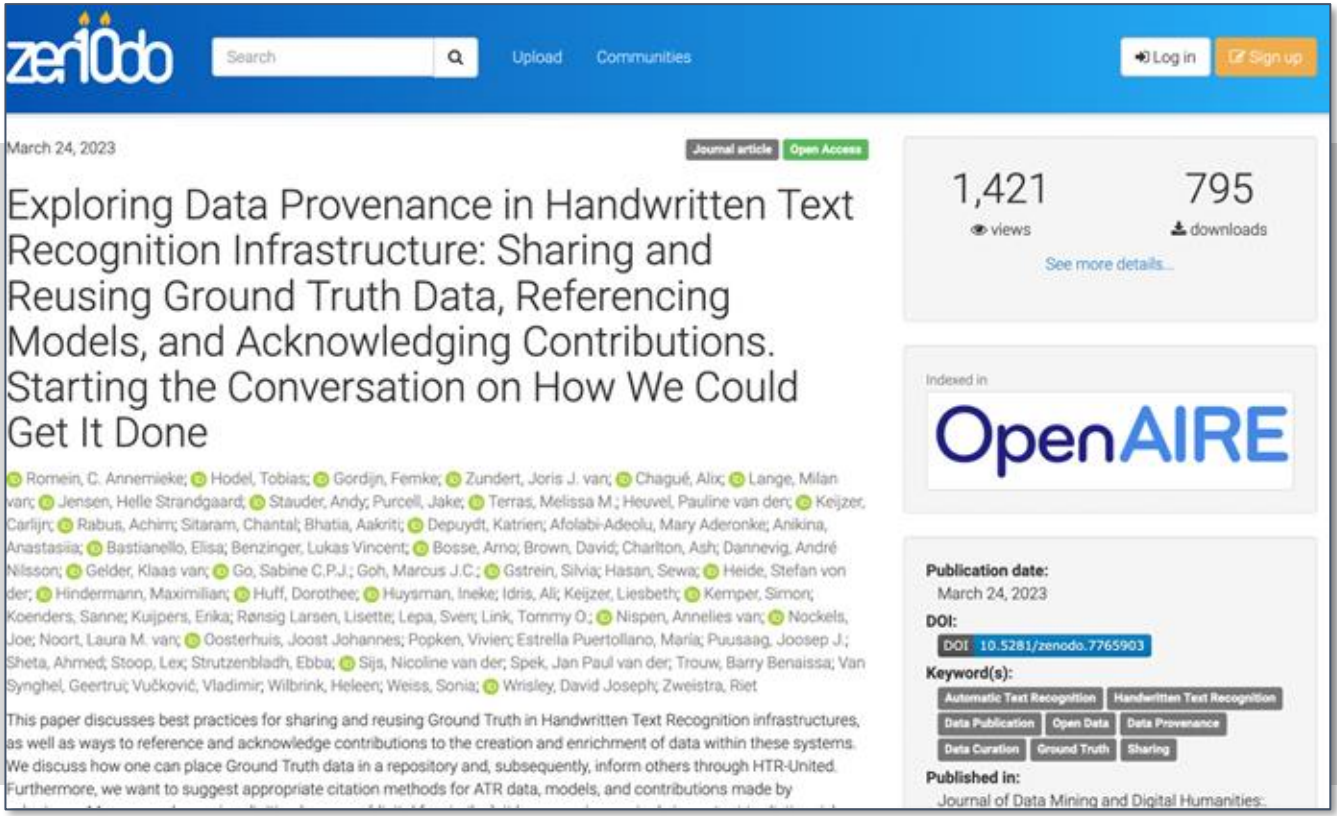


htr-united.github.io



Cooperate

Romein et al. 2024. "Exploring Data Provenance in Handwritten Text Recognition Infrastructure: Sharing and Reusing Ground Truth Data, Referencing Models, and Acknowledging Contributions. Starting the Conversation on How We Could Get It Done." *Journal of Data Mining & Digital Humanities*, 2024. https://doi.org/10.46298/jdmdh.10403.



Overview (Part II)

UNIVERSITÄT BERN

14:15	Test your own material
14:45	Show your results
15:05	Introduction to VLMs and command-line tools
15:25	Developing Workflows and Identifying Issues
15:45	Open Questions and Inputs on Demand

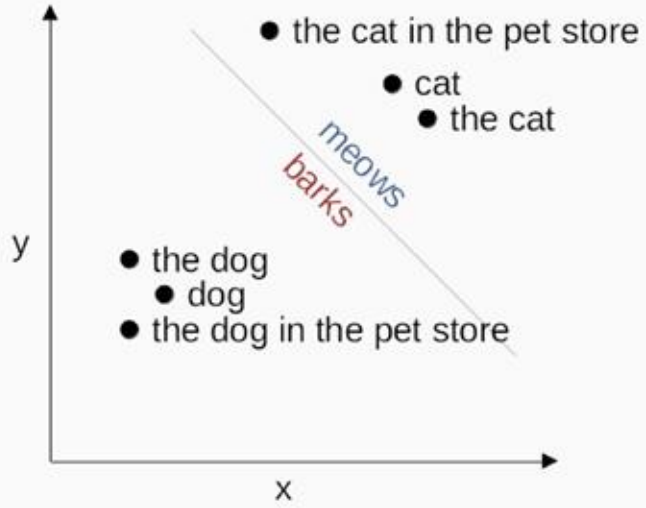
Test your own material and share it with us

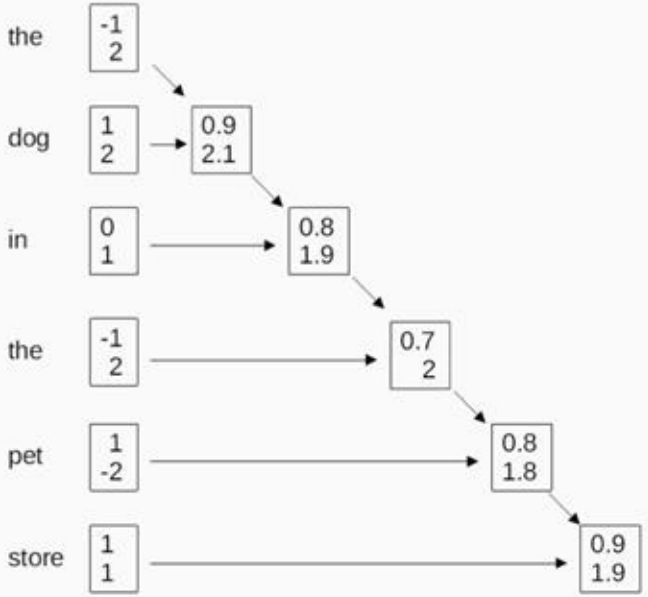
Introduction to VLMs

- How to LLMs work (way too short)
- Adding more dimensions (= multi-modal)

Vector Space

Representation as mathematical function/neural network





Quelle: Rico Sennrich, Interaktive Sprachmodelle: Ein Blick hinter die Kulissen.
https://kilof.unibe.ch/veranstaltungen/vergangene_veranstaltungen/kilof_tagung/index_ger.html

VLMs: Brainstorming potential challenges

Test out *commercial* VLMs

https://www.hodelweb.ch/vlm/

User: medievalist

Pw: kzoo2026!

Opportunity to compare

- Gemini
- QWEN-VL-Max
- Pixtral Large

Command line Tools

If you feel comfortable with the Command Line (and some Python)

- TrOCR:

https://github.com/microsoft/unilm/tree/master/trocr

- Kraken:

https://kraken.re/main/index.html

Workflows? Workflows!

UNIVERSITÄT BERN

Define:

- Goal / task
- Quality (threshold), be pragmatic
- A desired output

Identify:

- A suitable infrastructure
- Preexisting models
- An evaluation strategy