



HAL
open science

MEDUSA 0.1: Medieval European Documents Unified System for Automated text recognition System Report for the ICDAR 2026 Competition on Multilingual Medieval Handwritten Text Recognition

Théo Moins, Florian Cafiero, Jean-Baptiste Camps, Émilie Guidi, Lilla Conte, Brenna Hensley, Katarzyna Anna Kapitan, Carolina Macedo, Paola Peratello, Cecile Vermaas, et al.

► To cite this version:

Théo Moins, Florian Cafiero, Jean-Baptiste Camps, Émilie Guidi, Lilla Conte, et al.. MEDUSA 0.1: Medieval European Documents Unified System for Automated text recognition System Report for the ICDAR 2026 Competition on Multilingual Medieval Handwritten Text Recognition. 2026. <hal-05600991v3>

HAL Id: hal-05600991

<https://enc.hal.science/hal-05600991v3>

Preprint submitted on 30 Apr 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

MEDUSA 0.1: Medieval European Documents Unified System for Automated text recognition

System Report for the ICDAR 2026 Competition on Multilingual Medieval Handwritten Text Recognition

Théo Moins¹, Florian Cafero^{1,2}, Jean-Baptiste Camps¹, Lilla Conte¹, Emilie Guidi¹, Brenna Hensley¹, Katarzyna Kapitan¹, Carolina Macedo¹, Paola Peratello¹, Cecile Vermaas¹, and Chahan Vidal-Gorène¹

¹ École nationale des chartes – PSL, Paris, France

² EPITA, Paris, France

{abc,lncs}@chartes.psl.eu

This document is a system report prepared for the ICDAR 2026 Competition on Multilingual Medieval Handwritten Text Recognition (CMMHWR) and reflects the state of a system called MEDUSA (Medieval European Documents Unified System for Automated text recognition). It is made available online as a reference snapshot of the work at that stage. A revised and extended version, with additional training iterations and improved results, is forthcoming. Our approach combines a carefully assembled multilingual training corpus spanning more than twenty repositories with a multi-stage fine-tuning strategy built on a Vision Language Model (VLM). We first describe our data collection and curation pipeline, then detail the model architecture and training procedure, before reporting and discussing results on both our internal test set and the three official competition tasks.

Data

Following the competition’s philosophy of maximising robustness to unseen languages, we aimed to assemble the most diverse possible training, validation, and test sets in terms of languages, scripts, dates, and manuscript quality. A key challenge was that not all languages are represented by corpora of comparable size and diversity. Restricting ourselves strictly to data following CATMuS guidelines [17] would have limited our linguistic coverage. Furthermore, producing an original test set was necessary to enable fair comparison with models trained on the full CATMuS dataset (Kraken CATMuS 1.6.0 [18], Churro [22], etc.). Our work on the data part can be summarised as follows:

First, we surveyed more than twenty repositories containing European medieval HTR data (see Table 1), with a normalisation effort to convert everything to ALTO XML. Some repositories lacked line segmentation information, which was added by performing line segmentation, manual proofreading, and text alignment. This was complemented by the collection of text-only resources without associated images. This led to three data levels (see Figure 1):

1. *Platinum* (image/ALTO XML pairs following the target guidelines),
2. *Gold* (image/ALTO XML pairs following different guidelines),
3. and *Silver* (text only).

Each level is used in a specific way during training, in a curriculum that progresses from Silver to Platinum. Additional data augmentation was performed after initial results to reinforce under-represented languages, most notably Latin, which is one of the evaluation targets.

Second, we produced an original test set composed of 63 pages drawn from 36 documents across 13 languages, with a selection heuristic aimed at maximising diversity in terms of dates and scripts. Additional Platinum training data was also produced, including a total of 55 pages from 10 distinct Occitan documents, constituting a dataset in its own right [5].

Table 1. Description of datasets used to train our models.

Dataset	Ref.	Lang	Line or Page level	Gold or Platinum	No of lines
COMETA	[28]	Occitan	Page-level	Gold	118,105
Torino L-II-14	[6]	Old French	Page-level	Gold	36,823
FROC-MSS	[4]	Old French/Occitan	Page-level	Gold	3,636
iForal	[20]	Latin/Old Portuguese	Page-level	Gold	8,009
DISTINGUO	[3]	Latin	Page-level	Gold	15,190
AMSMB	[9]	Latin/Catalan	Page-level	Gold	3,369
Tridis	[24]	Multilingual	Line-level	Gold	166,784
HTR-School-Vienna-2025	[16]	Latin	Page-level	Gold	7,477
Paris Bible Project (PBP)	[29]	Latin	Page-level	Gold	1,606
St-Victor (M. Vernet) gold	[27]	Latin	Page-level	Gold	10,736
Wien ÖNB Cod. 2160 f. 164-184	[1]	Latin	Page-level	Gold	2,681
Bifrost	[13]	Old Norse	Page-level	Gold	873
Klosterneuburg	[2]	Middle High German	Page-level	Gold	4,758
Faithful transcriptions	[11]	Multilingual	Page-level	Gold	8,001
StABS Ratsbücher	[12]	Middle High German	Page-level	Gold	8,371
Inzigkofen	[10]	Middle High German	Page-level	Gold	8,321
Total (gold)					420,893
CMMHWR dataset	[8]	Multilingual	Page-level	Platinum	149,741
CATMuS medieval	[7]	Middle Dutch, Old English	Line-level	Platinum	47,084
GATMUZA	[5]	Occitan	Page-level	Platinum	2,117
TranscriboQuest 2025	[14]	Multilingual	Page-level	Platinum	1,278
HTR-School-Vienna-02	[26]	Old Czech	Page-level	Platinum	1,336
Padeřov-Bible	[15]	Old Czech	Page-level	Platinum	7,177
2024-medieval-czech	[19]	Old Czech	Page-level	Platinum	2,748
Total (platinum)					221,987
Original data		Multilingual	Page-level	Both	142,329
Total					642,880

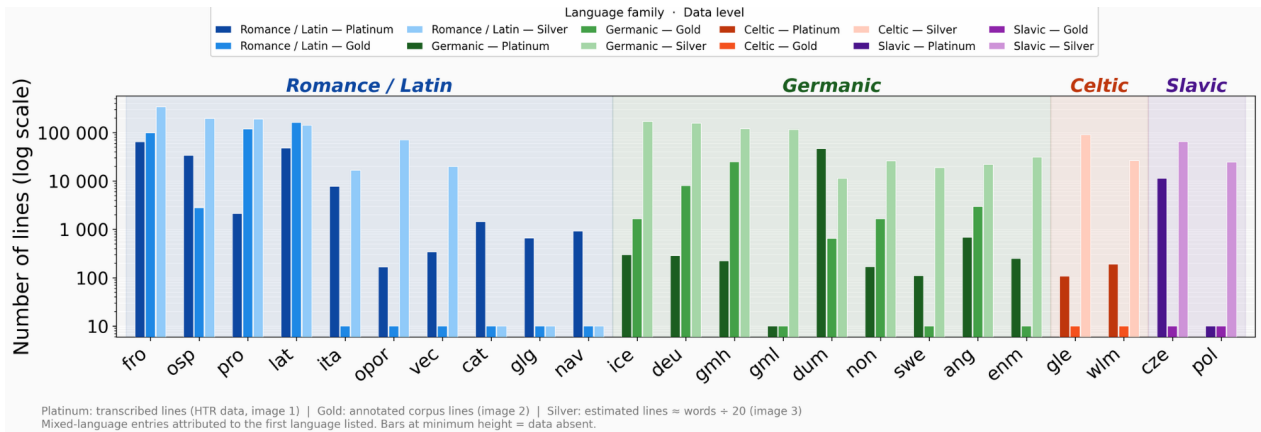


Fig. 1. Amount of data per language and level (platinum, gold and silver)

Model

Our system is based on fine-tuning a Vision Language Model (VLM), specifically Qwen 3.5 [21], in its 2B, 4B and 9B parameter variants, using LoRA fine-tuning as implemented by Unsloth [23]. This architecture offers two key advantages for our setting:

1. it leverages a model already proficient across many tasks and modern languages, so training focuses on adapting to new scripts and medieval languages rather than learning the transcription task from scratch,
2. it is naturally suited to multi-stage training : in the experiments reported here, we trained on a combination of Gold and Platinum data, with prompts adapted to dataset-specific conventions for the former, and targeting CATMuS guidelines for the latter.

Our best results were obtained by training the 4B and 9B variants on a mixed Gold and Platinum dataset over three epochs, plus three additional epoch of Platinum data only. The framework is further designed to accommodate a Silver pre-training stage, where text-only data would be used to first adapt the language component of the model; this stage, along with the generation of 500,000 synthetic lines for data augmentation, remains ongoing work and will be integrated in future iterations. The model’s output is further subjected to a post-processing step in which any character falling outside the CATMuS-defined character set is removed, ensuring compliance with the competition’s guidelines. Ultimately, the proposed framework could also serve for transcription under varying editorial conventions, parameterised according to abbreviation handling, allograph treatment, normalisation, and the use of u/v and i/j distinctions. While we have focused primarily on competition performance here, this direction will be explored further in future work.

Results

Following an extensive benchmark, **kraken-CATMuS 1.6.0** emerged as the strongest baseline on our internal test set and was therefore used as the primary point of comparison. Table 2 reports

Table 2. Unweighted average of per-language CER (%) and WER (%) on the internal and competition test sets. Best results per column are **bolded**.

Model	Params	Training data	Test set		Task 1		Task 2		Task 3	
			CER	WER	CER	WER	CER	WER	CER	WER
<i>Existing models (baselines)</i>										
Qwen 3.5	4B	-	50.8	93.4	-	-	-	-	-	-
Churro (<i>line-level use</i>) [22]	3B	CATMuS + other	38.1	80.1	-	-	-	-	-	-
TrOCR (medieval-base)	0.6B	CATMuS	27.7	70.2	-	-	-	-	-	-
Tridis [25]	8B	CATMuS + TRIDIS	21.4	60.9	-	-	-	-	-	-
kraken-CATMuS 1.6.0 [18]	4M	CATMuS	17.3	53.5	9.29	32.2	7.91	36.8	25.9	79.0
<i>MEDUSA fine-tuned models (ours)</i>										
MEDUSA-4B	4B	Gold/Platinum (3 ep.)	14.7	44.5	8.15	27.0	5.60	24.6	12.0	56.2
MEDUSA-4B	4B	Gold (2 ep.) + Platinum (2 ep.)	15.5	46.3	8.08	26.9	5.33	24.0	12.2	54.8
MEDUSA-4B	4B	Gold/Platinum (3 ep.) + Platinum (1 ep.)	15.1	43.5	8.03	26.7	5.20	22.2	11.5	54.6
MEDUSA-9B	9B	Gold/Platinum (3 ep.)	13.5	43.3	8.05	27.3	5.29	23.0	11.1	53.5
MEDUSA-9B	9B	Gold/Platinum (3 ep.) + Platinum (1 ep.)	13.2	42.6	8.03	27.1	5.24	22.2	10.8	52.6

Character Error Rate (CER) and Word Error Rate (WER) for all models evaluated on both our internal test set and the competition’s official test set.

Our best system (MEDUSA-9B) improves over the **kraken-CATMuS 1.6.0** baseline by 1.2% on Task 1, 2.9% on Task 2, and 15% on Task 3. We submitted an earlier iteration on Task 2 that marginally outperformed the 9B model by 0.2%, while the 9B variant yields the best results on task 3, and performs similarly to the 4B equivalent in task 1. Crucially, these results are achieved with a single, unified model architecture: unlike approaches that train task-specific systems, MEDUSA uses the same weights across all three tasks, with performance differences driven solely by the nature and distribution of the evaluation data rather than by any task-specific engineering. By looking at per-language scores, Figure 2 reveals that MEDUSA-9B’s most substantial gains over kraken-CATMuS 1.6.0 are concentrated in non-Romance language families, while also improving on Romance languages at the same time (Latin itself excluded). Reductions in CER are particularly marked for Germanic languages, where the baseline struggled most, and in the Slavic language provided in Task 3. Across the board, MEDUSA-9B improves upon the baseline in nearly every language tested, except Latin and Middle English, where our model’s advantage is more modest. A closer analysis of error patterns in these two languages will be the subject of dedicated future work.

Discussion

The competition provided a rigorous external evaluation framework for our system, confirming the viability of multi-stage VLM fine-tuning for multilingual medieval HTR. Analysis of results across languages highlights the benefit of curriculum training and data augmentation, particularly for Latin. Several limitations nevertheless remain and define our roadmap for future work:

page-level The current evaluation operates at the line level, thereby abstracting away the non-trivial challenges of page-level segmentation and reading-order prediction, which we plan to address in a subsequent iteration

guidelines Conduct a more systematic study of the impact of different transcription conventions on model performance,

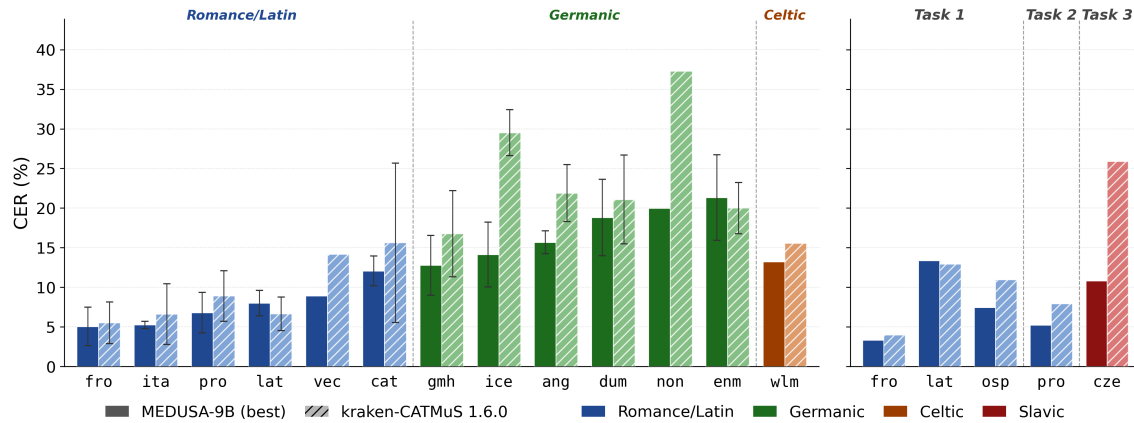


Fig. 2. Character Error Rate (CER, %) per language for MEDUSA-9B (best model, solid bars) and kraken-CATMuS 1.6.0 (hatched bars), grouped by language family. The left panel shows results on our internal test set (error bars indicate standard deviation across documents); the right panel shows results on the three competition test sets.

specialisation Explore specialisation by language family (Romance, Germanic, Latin, etc.).

Models are available on Huggingface³, along with the document analysis framework used for experiments on Github⁴. All original data produced will be the subject of a forthcoming publication, expected before the ICDAR conference in September.

Credits

Teams

- **Data:** Katarzyna Kapitan (KAK, lead), Jean-Baptiste Camps (JBC), Lilla Conte (LC), Emilie Guidi (EG), Carolina Macedo (CM), Paola Peratello (PP), Cecile Vermaas (CV)
- **Model:** Théo Moins (TM, lead), Brenna Hensley (BH), Chahan Vidal-Gorène (CVG), Florian Cafiero (FC)
- **Writing/Documentation:** Théo Moins (TM, lead), Jean-Baptiste Camps (JBC), Katarzyna Kapitan (KAK)

Roles summary

³ <https://huggingface.co/ENC-PSL/>

⁴ <https://github.com/TheoMoins/DocWorkflow>

Role	Description	Members
Conceptualization	Formulating research goals, questions, and overall project vision.	TM, JBC, KAK
Methodology	Designing research methods, workflows, or digital approaches.	TM, BH, JBC, KAK, CVG
Software	Developing or programming tools, scripts, or digital platforms.	TM, BH, JBC, CVG, EG, KAK, FC
Validation	Verifying results, ensuring accuracy of data and methods.	TM, BH, JBC, FC
Investigation	Conducting research, data collection, archival work, or fieldwork.	TM, JBC, KAK, LC, CVG, PP, EG, CM, CV
Resources	Providing materials, datasets, tools, or access to archives/collections.	TM, JBC, KAK, LC, CVG, PP, EG, CM, CV
Data Curation	Managing, cleaning, annotating, and preserving datasets.	TM, JBC, KAK, LC, CV, PP, EG, CM
Documentation	Preparing documentation.	TM, KAK, JBC
Writing – Original Draft	Preparing the initial version of text.	TM, JBC
Writing – Review & Editing	Revising, proofreading, and improving written content.	TM, FC, BH, CV, EG, CVG
Visualization	Creating charts, maps, or visual representations.	TM, EG
Supervision	Overseeing research activity, mentoring team members.	TM, KAK, JBC
Project Administration	Coordinating tasks, timelines, and team communication.	TM, KAK, JBC
Funding Acquisition	Securing financial support for the project.	JBC, KAK

Acknowledgements

Funded by the European Union (ERC, LostMA, 101117408). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

This work has received support under the Major Research Program of PSL Research University “CultureLab” launched by PSL Research University and implemented by ANR with the references ANR-10-IDEX-0001.

Ce travail a bénéficié d’une aide de l’État gérée par l’Agence Nationale de la Recherche au titre de France 2030 portant la référence « ANR-23-IACL-0008 »).

Biblissima+ bénéficie d’une aide de l’État gérée par l’ANR au titre du Programme d’investissements d’avenir intégré à France 2030, portant la référence ANR-21-ESRE-0005.

This work was granted access to the HPC resources of IDRIS under the allocation 2026-AD011015914R1 made by GENCI.

References

1. Attwood, Sweeney, Stitts, Audebrand, D’Amico, Geelhaar, Hofmann, Gnasso: Wien Önb cod. 2160 f. 164-184 ground truth from htr winter school 2022 (Dec 2022). <https://doi.org/10.5281/zenodo.7537204>, <https://doi.org/10.5281/zenodo.7537204>
2. Berger, M., Bolte, H., Führer, V., Hausleitner, F., Hutterer, S., Lüthi, T., Nancu, M., Passoni, E., Pataki, K., Schröcksnadel, S., Verri, G., Wegener, D., Hofert, S.: Klosterneuburg, stiftsbibl., cod. 48 - ground truth: Initial release (Dec 2022). <https://doi.org/10.5281/zenodo.7466928>, <https://doi.org/10.5281/zenodo.7466928>
3. Burghart, M., Yatsyk, S.: Distinguo : Ground truth for handwritten text recognition (htr) on collections of distinctions (late 13th to late 15th century) (2024). <https://doi.org/10.34847/NKL.48AD8B8D>, <https://nakala.fr/10.34847/nkl.48ad8b8d>
4. Camps, J.B.: FROC-MSS: Old French and Old Occitan Medieval Manuscripts HTR Data and Models. <https://github.com/Jean-Baptiste-Camps/FROC-MSS> (2018)
5. Camps, J.B.: Lo GAT MUZA: CATMUS-conformant HTR Ground-Truth Data for Medieval Occitan [Data set] (2026)
6. Camps, J.B., O’Connor, P.: Torino_L-II-14: HTR Training Dataset for the manuscript Turin, Biblioteca nazionale universitaria, MS L. II. 14 (2024), https://github.com/RESCAPE-Bibliissima/Torino_L-II-14
7. Clérice, T., Pinche, A., Vlachou-Efstathiou, M., Chagué, A., Camps, J.B., Gille-Levenson, M., Brisville-Fertin, O., Fischer, F., Gervers, M., Boutreux, A., Manton, A., Gabay, S., O’Connor, P., Haverals, W., Kestemont, M., Vandyck, C., Kiessling, B.: CATMuS Medieval: A multilingual large-scale cross-century dataset in Latin script for handwritten text recognition and beyond (Feb 2024), <https://inria.hal.science/hal-04453952>, working paper or preprint
8. Clérice, T., Kiessling, B.: Icdar 2026 competition on multilingual medieval handwriting recognition (Jan 2026). <https://doi.org/10.5281/zenodo.18270331>, <https://doi.org/10.5281/zenodo.18270331>
9. Coll Ardanuy, M., Cuadrada, C., Sarobe, R.: A Dataset for Handwritten Text Recognition in Medieval Notarial Charters Written on Parchment (2025), <https://dataverse.bsc.es/dataset.xhtml?persistentId=perma:BSC/OVBOMC>
10. Eichenberger, N.: Transcriptions from medieval manuscripts related to the augustinian canonesses in inzigkofen (Dec 2025). <https://doi.org/10.5281/zenodo.17978574>, <https://doi.org/10.5281/zenodo.17978574>
11. Eichenberger, N., Suwelack, H.: Faithful transcriptions data set: Tei/xml-encoded transcriptions of medieval theological manuscripts (Oct 2021). <https://doi.org/10.5281/zenodo.5582483>, <https://doi.org/10.5281/zenodo.5582483>
12. Hodel, T., Schoch, D., Dängeli, P.: Handwritten text recognition ground truth set: Stabs ratsbücher o10, urfehdenbuch x (Aug 2021). <https://doi.org/10.5281/zenodo.5153263>, <https://doi.org/10.5281/zenodo.5153263>
13. Kapitan, K.A., Vidal-Gorène, C.: Crossing the bifrost: Towards an open access fair htr model for old norse manuscripts (May 2025). <https://doi.org/10.5281/zenodo.15366896>, <https://doi.org/10.5281/zenodo.15366896>
14. McDonough, C., Henningsson, P., Denicolò, B., Bougrelle, R., Vieira, M.F., Salvo Garcia, I., Dehaut, C., Pinche, A.: Transcriboquest 2025 medieval vernacular religious texts (Sep 2025). <https://doi.org/10.5281/zenodo.17062963>, <https://doi.org/10.5281/zenodo.17062963>
15. Michalcová, A., Bazelides, K., Hajič, J., Pěnkavová, E., Maniaková, L., Kreisingerová, H., Filipová, J., hung Lu, C., Dvořáková, M.: Padeřov-bible-handwriting-ground-truth: Initial release (Dec 2022). <https://doi.org/10.5281/zenodo.7467034>, <https://doi.org/10.5281/zenodo.7467034>
16. Odstrčilík, J., Čermáková Lukšová, Z., Kovács, A., Andre, F., Barak, H., Bold, N., Cassi, R., Coppola, A., Kalashnikova, O., Kostoff-Kaard, J., Lefle, N., Leoni, L., Noworyta, M.T., Paulsen, A., Rossi, F., Šámal, P., Ugliano, D., Valodzina, A., Warren, T.P., Zachara-Związek, U., Ziemińska, A.: HTR Winter School in Vienna 2025 – Late Medieval Latin Group: Ground Truth Dataset for Late-Medieval Latin

- Scripts (2025), created during the Winter School: HTR of Historical Sources, Vienna, 2025. Licence CC-BY 4.0. Consulté le 2 avril 2026
17. Pinche, A., Clérice, T., Chagué, A., Camps, J.B., Vlachou-Efstathiou, M., Levenson, M.G., Brisville-Fertin, O., Boschetti, F., Fischer, F., Gervers, M., et al.: Catmus-medieval: Consistent approaches to transcribing manuscripts. In: Digital Humanities-DH2024 (2024)
 18. Pinche, A., Clérice, T., Vlachou-Efstathiou, M., Chagué, A., Camps, J.B., Gille Levenson, M., Brisville-Fertin, O., Boschetti, F., Fischer, F., Gervers, M., Boutreux, A., Manton, A., Gabay, S., Ferrante, G.: Catmus medieval (Mar 2025). <https://doi.org/10.5281/zenodo.15030337>, <https://doi.org/10.5281/zenodo.15030337>
 19. Plechatý, M., Steiner, V., Švarc, J., Marie, H., Michalcová, A., Hejdová, T., Spěváčková, M., Katscher, D., Škvrňák, J.: HTR Winter School 2024 - Medieval Czech - Prague Bible (1488) (Dec 2024)
 20. Projet iForal: iForal Dataset: Medieval Portuguese Manuscripts HTR Data. <https://github.com/Arch-W/iForal-Dataset>
 21. Qwen Team: Qwen3.5: Towards native multimodal agents (February 2026), <https://qwen.ai/blog?id=qwen3.5>
 22. Semnani, S., Zhang, H., He, X., Tekgürler, M., Lam, M.: Churro: Making history readable with an open-weight large vision-language model for high-accuracy, low-cost historical text recognition. In: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. pp. 34765–34812 (2025)
 23. Singhapoo, K., Inthanil, A., Pillai, A.: Fine-tuning ai models with limited resources. In: 2025 11th International Conference on Engineering, Applied Sciences, and Technology (ICEAST). pp. 148–151. IEEE (2025)
 24. Torres, S.: Tridis (revision e8d811f) (2025). <https://doi.org/10.57967/hf/5001>, <https://huggingface.co/datasets/magisttermilitum/Tridis>
 25. Torres Aguilar, S.: Dual-style transcription of historical manuscripts based on multimodal small language models with switchable adapters (2025), <https://hal.science/hal-04983305>, working paper or preprint
 26. Veličkaitė, V., Michalcová, A., Landová, L., Reimannová, M., Kostelníková, D., Kostelník, A., Urbánková, M., Jakubcová, V., Hejdová, T., Javoříková, A., Pytlíková, M., Kozlovská, K., Dušek, J., Doušková, B., Švecová, K., Nová, V., Filipová, J., Soukupová, V., Pěnkavová, E., Katscher, D., Spěváčková, M.: HTR Winter School 2025 - Medieval Czech - Biblioteka Jagiellonska BJ Rkp 441 IV (Dec 2025)
 27. Vernet, M.: Saint victor ms dataset (abbreviated and expanded alto) (Jan 2023). <https://doi.org/10.5281/zenodo.7510410>, <https://doi.org/10.5281/zenodo.7510410>
 28. Wiedner, M.: COMETA : Corpus de l'occitan médiéval comparatif et annoté: Provence et Languedoc, <https://zenodo.org/records/15300719>
 29. Wrisley, D., Project, T.P.B., Gueville, E.: parisbible/ground_truth: Ground truth v1.0.0 for the paris bible project (Feb 2023). <https://doi.org/10.5281/zenodo.7653691>, <https://doi.org/10.5281/zenodo.7653691>