

# Efficient OCR Post-Correction and Annotation for Neo-Latin Corpora

A Human-in-the-Loop Workflow Integrating LLMs in  
Digital Scholarly Editing

---

Sarah A. Lang

Max Planck Institute for the History of Science (Berlin)

Aarhus/online, 2025-09-24

# Bridging Computational and Humanities Needs

This presentation does not aim to provide research novel for digital humanities.

Rather, it seeks to bridge digital methods with the actual needs of historians and humanities scholars.

As computational approaches move closer to computer science, we should not forget about the more basic uses of digital methods that are more closely aligned with what scholars actually need.

# Large Scale DH vs. Humanists Expectations

## High Accuracy Standards for Editions

- **Minimal Tolerance for error** digital scholarly editing.
- Lesser quality can render data unusable or not taken seriously by scholars.
- **Implication:** Perfectly corrected data is valuable but resource-intensive. Can AI help?

## Starting Points: NOSCEMUS ERC & NOSCEMUS GM

- **Corpus:** Representative for Neo-Latin scientific literature.
- **HTR Model:** The *NOSCEMUS GM* on Transkribus.
- **Access:** Transkribus is now a (mostly) closed service.
- **Criticisms:** TEI-XML data based on imperfect OCR is met with resistance by scholars.

# A Human-in-the-Loop Post-Correction Workflow

## Pressing Question

*How can current technologies assist in creating perfectly corrected editions more efficiently?*

## Suggested Solution

Corrections suggested by LLMs are manually reviewed using visual diff tools. This offers an efficient yet fully human-controlled process.

## Approach

- Use the strong *NOSCEMUS GM* as base OCR/HTR.
- Employ LLMs to suggest corrections. → this part can be automated via API.
- Integrate visual diff tools (e.g. colordiff or eScriptorium) for manual verification.
- → **Manual correction remains central for accuracy and conformance to guidelines** because LLMs often degrade OCR quality (Boros et al. 2024). They tend to ignore or mix-up transcription standards, making outputs unusable for OCR re-training / generating ground truth



# Leveraging colordiff (1)

106

## Hieroglyphicorum Lib. 3.

& virilia in mare, vnde Venus exorta est: Huic victoriae contra Titanes & patrem Saturnum Iovi caeteri dii adfuerunt: Apollo cecinit cythara eius victoriam purpurea togâ decorus & coronatus lauro: Tum ferunt Iovem Vestae proposuisse, per quam adeptus erat imperium, ut quod mallet, eligeret, at hanc virginitatem poposcisse & primitias sacrificiorum. Fama est Bacchum quoque bello, contra Titanes gesto disceptum fuisse, cuius cor palpitans Pallas ad Iovem detulit; Post gigantes contra Iovem surrexerunt, quos ille fulmine disiecit & inter eos Aegeonem, (cui 100. manus erant, capita 50.) Iovem regno depellere conantem sub Aetnam montem detrulit; Vnde quoties latus mouet, Aetna flammâs eructat: Eodem partu cum Iove Iuno nata dicitur, ab Horis educata; Vel ab Oceano & Tethye, ut alijs placet: Huius potiundae desyderio captus Iupiter sese in cuculum mutauit, quicum tempestate oborta tremere frigeretque in gremium Iunonis aduolauit: qua veste tectus, eam resumptâ primâ formâ compressit & post in uxorem duxit; Ex Iunone & Iove nascitur Mars, Arge, Ilithya & Hebe: Ab alijs his additur Vulcanus; Quem Lucianus subuentaneo conceptu è Iunone absque maris congressu natum fingit: Credita fuit esse diuitiarum dea ab antiquis; Eamque fama est Iouem aliquando alie suspendisse, pedibusque illius duas incudes appendisse atque aureum vinculum iniecisse manibus; Cumque ita Iuno penderet in aethere, caeteros deos id graviter tulisse, neque soluere tamen potuisse: Attribuuntur ei quatuordecim nymphae; Pavoque facer habetur; Quia propter illam caesus fuit Argus a Mercurio & in illam avem mutatus: Anser alias fuit auis sacrata Iunoni; Sic & vacca ex animalibus; Quia ex figuris Hieroglyphicis Aegypti est, Iunonem denotans: De fabulis Iunonis, ut ante sub Saturno, Natalis dicit chemicos artifices conatos esse nonnullas partes ad suos ignes & ad sua vascula retrahere: Dicitur Iuno (inquiunt) Saturni & Opes filia, Iovisq; soror & coniunx, & ante Iovem nata eodem partu, Reginaq; esse deo-

Iuno.

Entellus in-  
stat.

106 -1,43 +1,43 00

106

## Hieroglyphicorum Lib. 3.

& virilia in mare, vnde Venus exorta est: Huic victoriae contra Titanes & patrem Saturnum Iovi caeteri dii adfuerunt: Apollo cecinit cythara eius victoriam purpurea togâ decorus & coronatus lauro: Tum ferunt Iovem Vestae proposuisse, per quam adeptus erat imperium, ut quod mallet, eligeret, at hanc virginitatem poposcisse & primitias sacrificiorum. Fama est Bacchum quoque bello, contra Titanes gesto disceptum fuisse, cuius cor palpitans Pallas ad Iovem detulit; Post gigantes contra Iovem surrexerunt, quos ille fulmine disiecit & inter eos Aegeonem, (cui 100. manus erant, capita 50.) Iovem regno depellere conantem sub Aetnam montem detrulit; Vnde quoties latus mouet, Aetna flammâs eructat. Eodem partu cum Iove Iuno nata dicitur, ab Horis educata; Vel ab Oceano & Tethye, ut alijs placet: Huius potiundae desyderio captus Iupiter sese in cuculum mutauit, quicum tempestate oborta tremere frigeretque in gremium Iunonis aduolauit, a qua veste tectus, eam resumptâ primâ formâ compressit & post in uxorem duxit; Ex Iunone & Iove nascitur Mars, Arge, Ilithya & Hebe. Ab alijs his additur Vulcanus; Quem Lucianus subuentaneo conceptu è Iunone absque maris congressu natum fingit: Credita fuit esse diuitiarum dea ab antiquis; Eamque fama est Iouem aliquando alie suspendisse, pedibusque illius duas incudes appendisse atque aureum vinculum iniecisse manibus; Cumque ita Iuno penderet in aethere, caeteros deos id graviter tulisse, neque soluere tamen potuisse: Attribuuntur ei quatuordecim nymphae, Pavoque facer habetur; Quia propter illam caesus fuit Argus a Mercurio & in illam avem mutatus: Anser alias fuit auis sacrata Iunoni; Sic & vacca ex animalibus; Quia ex figuris Hieroglyphicis Aegypti est, Iunonem denotans: De fabulis Iunonis, ut ante sub Saturno, Natalis dicit chemicos artifices conatos esse nonnullas partes ad suos ignes & ad sua vascula retrahere:

## Leveraging colordiff (2)

## Version control powered comparison (git colordiff)




```
git diff --color-words=. transkribus-output/0122_arcana.txt gpt-output/0122_arcana.txt:
```

```
git diff --word-diff=color --word-diff-regex="[ ]+|^]+" transkribus-output/0122_arcana.txt gpt-output/0122_arcana.t
```

[illegible]

Model: *gpt-3.5-turbo* via API; script: <https://github.com/sarahalang/LLM-powered-OCR-correction/blob/main/korr-ocr.py>


# Diff functionality in eScriptorium

 Line #12

First of all, here is what I told, with the m

## First of all, here is what I told, with the mo

by (eScriptorium) on Mon Feb 27 2023 12:17:48 GMT+0100

 [Toggle transcription comparison](#)

First of all, here is wWhat I told, wwith the most deep

First of all, here is what I told, with the most deep

First of all, here is whvnat I told, wwith the most deep.

# Making things more difficult

## Challenge: Special Characters

mixed-up column problem has been fixed in Transkribus in the meantime

Transkribus Lite 1687\_Magnet\_Pharmacopoea-Schr... View Text Layout 20.9.2022, 11:15 1 150

℥. Calc. ☉ fulminantem, redige in flores per peculiare aliquod instrumentum argenteum deauratum: Ex floribus extrahe tinctur. cum spiritu mannæ, abstrahere ad consistentiam pro libitu.

Vires. Sudorem, cumque fœdissimum pellis, felicissimè morborum quorumcunque malignorum & venenatorum seminarium secum vehens. Cl. Helv. Dietrich.

Dosis guttæ aliquot.

12. Aurum potabile cum -u Sal. communis Tartari.

F. āāā. ʒj. cum ʒvi. ♀ more vulg. superfluum ♀ ium per corium exprime, globulum restantem tere cum ana Sal. Com fusi, fac evaporare ♀ ium mediocri Δ, salem ablue aq. calidā, calcem ☉ tere fortiter cum ʒiiij. ʒ communis, pone in regulā fulminatoria (cine Treibschergen.) Un Creuset nommé

Dosis gran. 2.

1. Calcina c  
2. Reverber  
vis instar boli a  
3. Solve cum  
bando multoties  
hendo.  
4. Solve in spirit.  
(3.) hinc abstrahere. B  
17.

℥ Auri fulminantis ☉-i volatilis part. i cum, in quo salis pro sed tale menstruum n gerantur in B. tepenti

3. Solve cum V rectificatiss. digerendo ac cohobando multoties, (20.) & ultra: dein coagula abstrahere.

heudo.

Dosis guttæ aliquot.

4. Solve in spirit. Sal. com. digere dies aliquos

12. Aurum potabile cum -u Sal.

(3.) hinc abstrahere. Billich. in Exerc. th. 101.

communis Tartari.

17. Tinctura Oi.

F. 2ia. 3j. cum a VI. & more vulg. superfluum Fium per corium exprime, globulum restantem tere + Auri fulminantis probè edulcat. pat. I. shi cum ana Sal. Com fusi, fac evaporare Q ium mediocri tne & i volatilis part. IV. sunt qui Oleum & cauli-

# Case Study: Recognising Alchemical Symbols

## The Project

- **Actual Research Goal:** Enable automatic recipe extraction – but the need for manual PCR correction remains a major bottleneck.
- **Aim:** Retrain model for rare alchemical symbols.
- **Corpus:** Complex page layouts from Magnet's *Pharmacopoeia* = Johannes Jacobus Magnet, *Pharmacopoea Schrödero-Hoffmanniana illustrata et aucta etc* (1687). URL: <http://data.onb.ac.at/rep/10542338>.
- **High correction time** (30–60 minutes per page); pages rich in symbols are most valuable but most labor-intensive.

## Crowdsourcing by students does not work

- Requires Latin proficiency.
- Needs familiarity with abbreviations and OCR error types.
- Demands high accuracy in symbol identification and adherence to transcription guidelines.

# Challenges with alchemical special characters

d ʒj. & ʒj. per se vel cum aliis medica-

Magisterium Viperarum.

rum seu serpent. nostr. & in locis sanis  
putridis degentium q. v.

ous & caudis detruncatis, exemptisque  
in cordibus, hepatibus, ac ipsa spina,  
confracta, ) in ☉ l. q. immerge.

tres dies, postea ad perfumandum & ex-  
libro puro imposito, subtus suffimento,  
excitato ex bacc. juniperi p. iij. lauri p.

finem adjicitur succini ʒj. asse dulcis ʒʒ.

erpentes ita perfumatos, in olla nova vi-

per fornacem, vel & in sole exsiccantur,

em redigi possint. Jam  
pulveris q. v. ☉ is ammoniaci l. q. ad su-

scil. ij. vel iij. digitorum.

ecantatum, & prius filtratum, seu cola-

ur ʒʒi ex ☉ salt. q. l. præcipitatio-  
lutio & edulcoratio cum aqua simplici,

*citri & angelica.*

Diaphoreticum compositum Dn. D. Cn.

℞. ʒii fixi diaphoretici ʒij. bezoari. mineralis  
ʒiv. corall. rubr. preparator. ʒj. pulv. viperin. ex cord.  
& hepat. croci nostri solaris ana ʒj.

Misceantur, fiatque ex his omnibus, in mortario vi-  
treo pulvis subtilissimus, cui adhuc admisceatur  
essent. citri ex cort. exter. rec. extr. gr. xv.

servetur pro usu in vitro benè clauso.

in hun finem excitato ex hace. juniperi q. iij. lauri .  
j. ubi & cirea finem adicitur succini ʒj. asse dulcis ʒj.

R. Aeras, Vita, utus, ubi q. j. si.

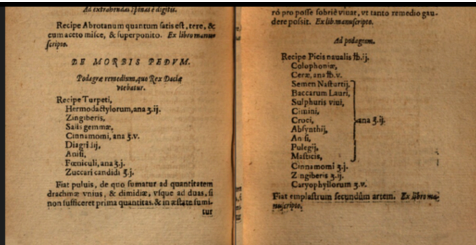
tracus Sijj. Sepentes ita perfumatos, in olla nova vi-  
ana ʒj.

trata, adhuc super frnacem, vel & in sole exsiccantur,

Misceantus pucum onis rita ʒ ʒ. plis,  
ut in pulvenem redigi possint. Fam  
subtilissimus.

R. Hujus pulqueris q. a. is amomiaci que ad su-  
vel q. Si dipintis fri, Pulei, ugris aliis  
pereminentiam scill. ij. vel iij. digitorum.

# Deciphering alchemical symbols with LLMs



Can you transcribe the text in this facsimile of a historical Latin recipe book of secrets? Make sure to get the columns right and also, transcribe historical symbols correctly as unicode characters. Use the layout and your understanding of the content to detect single recipes.



## Books of Secrets Recipe Extraction

To perform a faithful transcription of the text in the image provided, I will begin addressing the left-hand page (labeled as "186"), followed by the right-hand page (labeled as "187"), adhering to the column structure and employing Unicode characters where applicable to represent historical symbols. Please note that quality of the image may affect the accuracy of transcription.

## Books of Secrets Recipe Extraction

Left page (186):

Copy code

### DE SECRETIS

Fellis porcini 3j.  
Cerae quantum sufficit est.  
Misce, & fiat unguentum, quo unguantur vertices, & siliquae morbi, & Roscellus.

### DE MORBIS MANUUM.

Ad manus tremorem.  
Recipe Abrotanum quantum satis est, tere, & cum aceto nitrico, & superponito. Ex libro manuscripto.

### DE MORBIS PEDUM.

Podagrae remedium, quod Rex Dacie utebatur.  
Recipe Turpeti,  
Hemadactylorum, ana 3j.ii.  
Zingiberis,  
Salis gemmae,  
Cinnamomi, ana 3 v.  
Diagiri,  
Anisi,  
Feniculi, ana 3 j.  
Zuccari candidi 3j.

→ LLMs have been shown to be inefficient at OCR post-correction, usually degrading results (Boros et al. 2024)

# Can the LLM do it with a more simple task?

## An attempt at automation..

You're an expert academic OCR corrector for historical text. You do this by comparing a supplied initial OCR/HTR output with the original image from which it was transcribed. The goal is to produce a perfect diplomatic transcription, following the transcription rules of the original document as evident from the provided base transcription. These rules include silently expanding certain abbreviations, normalizing *i/j* and *u/v*, and using the logical negation sign (*~*) to indicate a line break within a word. Additionally, it removes OCR artifacts or junk characters from faulty OCR transcription. The output must retain the original text structure (e.g., line breaks) to support version control visualization. The text may contain multiple languages, predominantly Latin, and special characters (e.g., *R*, *3*, *lb*, and alchemical symbols) should be accurately represented using their Unicode values. If unsure about a special character, represent it as '@'.

I'm sending one example page with manually corrected text that you can use as ground truth to understand the desired output (page 721). Your task is to take the draft transcription for p714 and improve it (correcting OCR errors and correctly representing any historical symbols as their unicode equivalents).

Output only the corrected text.



magnet-p714-output-transkribusm  
Document




magnet-p721-haendischKorrigiert.L  
Document

You're an expert academic OCR corrector for historical text. You do this by comparing a supplied initial OCR/HTR output with the original image from which it was transcribed. The goal is to produce a perfect diplomatic transcription, following the transcription rules of the original document as evident from the provided base transcription. These rules include silently expanding certain abbreviations, normalizing *i/j* and *u/v*, and using the logical negation sign (*~*) to indicate a line break within a word. Additionally, it removes OCR artifacts or junk characters from faulty OCR transcription. The output must retain the original text structure (e.g., line breaks) to support version control visualization. The text may contain multiple languages, predominantly Latin, and special characters (e.g., *R*, *3*, *lb*, and alchemical symbols) should be accurately represented using their Unicode values. If unsure about a special character, represent it as '@'.

I'm sending one example page with manually corrected text that you can use as ground truth to understand the desired output (page 721). Your task is to take the draft transcription for p714 and improve it (correcting OCR errors and correctly representing any historical symbols as their unicode equivalents).  
Output only the corrected text.

Now, I'll correct the OCR output for p714 based on the reference style from p721. This involves:




1. **Fixing OCR errors** (e.g., "atescendum supendum" → "arescendum suspensendum").
2. **Expanding abbreviations silently** where evident.
3. **Correcting character recognition** (e.g., *R*, *3*, *lb*, etc.).
4. **Retaining historical formatting**, including word breaks with  where applicable.

Message ChatGPT



# Example Prompt for LLM-Assisted OCR Correction

## Prompt

“ You’re an expert academic OCR corrector for historical text. You do this by comparing a supplied initial OCR/HTR output with the original image from which it was transcribed. The goal is to produce a perfect diplomatic transcription, following the transcription rules of the original document as evident from the provided base transcription. These rules include silently expanding certain abbreviations, normalizing i/j and u/v, and using the logical negation sign (¬) to indicate a line break within a word. Additionally, it removes OCR artifacts or junk characters from faulty OCR transcription. The output must retain the original text structure (e.g., line breaks) to support version control visualization. The text may contain multiple languages, predominantly Latin, and special characters (e.g., , , , and alchemical symbols) should be accurately represented using their Unicode values. If unsure about a special character, represent it as '@'. Output only the corrected text. ”

## Resources

- <https://github.com/sarahalang/LLM-powered-OCR-correction>
- Example images and texts in directory: *DHd2025-workshop-symbolle-postcorrection* 13/19
- <https://www.transkribus.org/en/model/print-latin-texts-15-th-19th-htr>

## The results

## Conclusions?

```
git diff --color-words=. magnet-p714-output-transkribusmodell.txt magnet-p714-gpt4o-corrected.txt:
```


```
git diff --word-diff=color --word-diff-regex="[ ]+|^+" magnet-p714-output-transkribusmodell.txt magnet-p714-gpt4o-corrected.txt:
```

peracto, serpentum sem **vinetorum uiperarum**  
in **vinetorum** fructula, & ad **pustulandum**  
**pustula**, **pustula**,... suffimento subtus ita exo-  
tarum uniperi iij, sulci quigar. iij.  
a & bent prius excicata, simul cum spina-  
edigantur in puluerem subtilissimum, qui pu-  
in vitro puto **incolu**, **incolu**, in loco si-  
est. Signetur puluis uiper arum simplex.  
hepata ae lingua simili modo in alcool  
vel in  $\equiv$  juniperinum imposita asservetur  
a usum, pro **chieria**, **theria**, **ostrea**, **ostrea**  
iijus pulueris. Contra omnia venena. in-

magnet-p714-gpt4o-corrected.txt ×

R. Viperarum sem fergent. usti, & in locis fanis  
nontusti & non purridis Argentum q. v.

Fusque capitibus & caudis detruncatis, exemptisque eo pulis  
subrilissimus, cui adlus admeceaur.  
interant, cum cordibus, hepatibus, ac ipsa spina,  
culero prius confrica, in aci s. q. in troce.  
Stent ita per tres dies, postes ad perfundum & ex.  
siccandum, cribro poro imposito, subrus suffimento,  
in hun finem excitato ex haet. juniperi P. iij. lauri q.  
j. ubi & cirta finem addicitur succini ij. ae dulcis ij.  
storacis, siij. Serpentes ita perfumato, in olla nova vi-  
trata, adhuc super radices, vel & in sole exsiccantur,  
ut in pulverem rediri possint. Fam.

. Hujus pulqueris q. . is ammoniaci q. ad su-  
pereminentiam scill. ij. vel ij. di.  orum.

in solum, decantarum, & prius filtratum, seu colatum, infunditur dei i ex e fact. q. s. praecipitatio. ne peracta, ablutio & decoctio cum aqua simplici, seu scorgoneræ, instituitur. Pulvis tandem in charta. emporiti relictus exsiccat & ad usum reservatur.

Vires & Dusi conteniunt, cum viperarum pulvetes,  
superius simplici modo parato.

**Pulvis viperinus compositus.**

Magisterium Viperarum.

℞. Viperarum seu serpent. nostr. & in locis sanis  
mo. rosts & non putridis degentium q. v.

Esque capitibus & caudis detruncatis, exemptisque  
intercanis, cum cordibus, hepatibus, ac ipsa spina,  
(cultro prius constricta,) in ☉ i. q. immerge  
Stent ita per tres dies, postea ad perfundandum & ex-  
siccandum, cribro puro imposito, subtus si iussimero,  
in hunc finem excitato *ex bac. juniperi* ℥ ij. *lauri* p.  
j. ubi & circa finem adijciat *succini* ℥j. *asse dulcis* ℥b.  
*styracis* ℥ij. Serpentes ita perfumatos, in olla nova vi-  
treata, adhuc super fornacem, vel & in sole exsiccantur,  
ut in pulverem redigi possint. Jam

~~24. *Huius pulueris q.v. n. & is ammoniac f. q. ad su-*  
pereminentiam scil. ij. vel iij. digitorum.  
In solatium, decantatum, & prius filtratum, seu cola-  
tum, infunditur *q. q. ex & f. q. f. precipitati-*  
one peracta, ablutio & edulcoratio cum aqua simpliciter  
seu scorzonera iustificatur. Pulvis tandem in charta~~

magnet-p714-output-transkribusmodell.txt

R. Viperarum sem fergent. usti, & in locis fanis  
nontusti & non purridis Argentum q. v.

Fusque capitibus & caudis detruncatis, exemptisque co pul  
subrillissimus, cui adlus admeisaur.  
interantis, cum cordibus, hepatibus, ac ipa spina,  
luto prius contracto, ac s. q. immerge.  
Stent ita per tres dies, postes ad perfundum & ex.  
siccandum, cribro puro imposito, subrus suffundito,  
in hun finem excitato ex haet. juniperi P. iij. lauri q.  
j. ubi & cirta finem adicitur succini j. ae dulcis j.  
storacis. Sij. Serpentes ita perfumato, in olla nova vi.  
trata, adhuc super frumacel, vel in sole exsiccantur,  
ut in pulvenem redire possint. Fam.

. Hujus pulqueris q. . is ammoniaci q. ad su-  
pereminantiam illi. ii. vel ii. digitorum.

in solutum, decantarum, & prius filtratum, seu colatum, infunditur dei i ex e fact. q. s. praecipitatio. ne peracta, ablutio & dulcoratio cum aqua simplici, seu spongiferae, instituitur. Pulvis tandem in charta. empori relictus exsiccat & ad usum teserva- Vires & Dusi contiunt, cum viperarum pulvetes, superius simplici modo parato.

Pulvis viperinus compositus.

# Why prefer my Editor-in-the-Loop Workflow?

## Human-Controlled Correction

- LLM-generated suggestions are integrated manually.
- Visualisation supports efficient incorporation of edits.
- Editors retain full control over the final output.
- No risk of hallucinations, even with special characters.
- All output is verified by humans before finalisation.

## Benefits

- LLM-assisted, but avoids typical LLM pitfalls.
- Speeds up correction from high-quality HTR to edition-level accuracy – cuts time roughly in half.
- LLMs identify subtle issues that humans may overlook (e.g., 'l' vs. 'I', missing spaces, punctuation).

# Current Limits of LLM-powered OCR Post-Correction

## Challenges

- The task is highly specialised and hard to outsource.
- LLMs fail due to lack of exposure to rare material and non-standard knowledge. They normalise data, ignore guidelines, and degrade OCR output.
- Even with Unicode input, LLMs hallucinate corrections.

## Current Findings

- Traditional machine learning is necessary for progress in this area.
- Some success with more frequent measurement symbols.
- Alchemical symbols are too rare for effective training on only single pages.

# The Dangers of Hallucinations

## Arthur Dee and the fake preface

- LLM generated a plausible early modern preface.
- None of it matched the actual manuscript.
- Hallucinations are especially dangerous due to their believability / plausibility.
- Even when prompted with transcription segments, models fabricate.

## Conclusion

This reinforces the necessity of a human-in-the-loop approach that leverages LLMs to help generate further training / fine-tuning data to feed into traditional ML models (more reliable for our tasks).

# Conclusions

## Critical Reflection

- Does AI reduce effort, or leave only tedious tasks?
- Historians risk being reduced to AI fact-checkers of hard-to-check, highly plausible looking content.
- Current state-of-the-art studies (e.g. Boros et al. 2024) question the efficiency of such workflows.

## AI vs. Humanities scholars

- Historians' skills remain indispensable.
- AI can assist but not replace human judgement in critical tasks.
- The human-in-the-loop approach remains simple yet effective.

## References

---

- [1] Emanuela Boros et al. **“Post-Correction of Historical Text Transcripts with Large Language Models: An Exploratory Study”**. In: *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)*. Ed. by Yuri Bizzoni et al. St. Julians, Malta: Association for Computational Linguistics, Mar. 2024, pp. 133–159. URL: <https://aclanthology.org/2024.latechclfl-1.14/>.