

Transcribing Machines:

HTR as a Problem of Critical AI

David Joseph Wrisley, NYU Abu Dhabi
Institute Lecture, DHSI, 12 June 2026

جامعة نيويورك أبوظبي



NYU | ABU DHABI

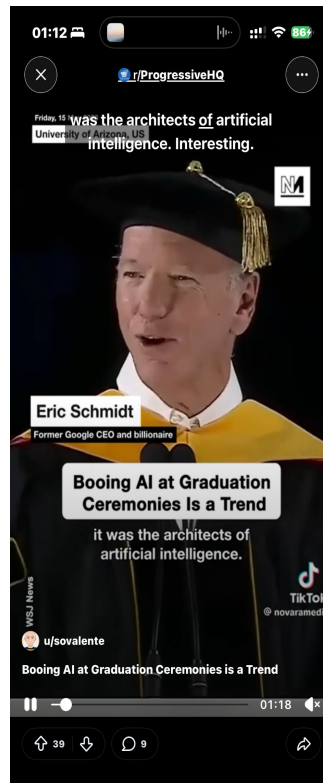
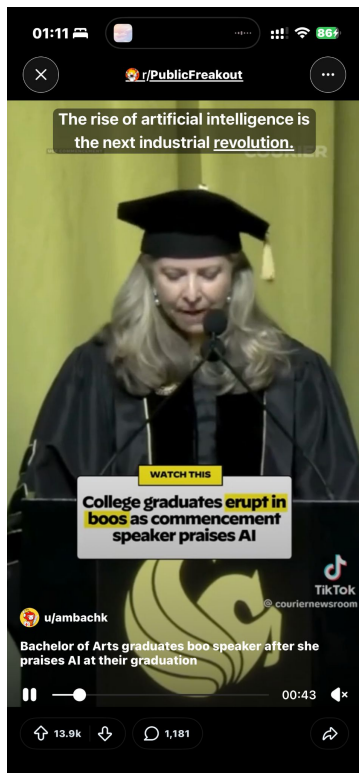


tinyurl.com/transcribingmachines

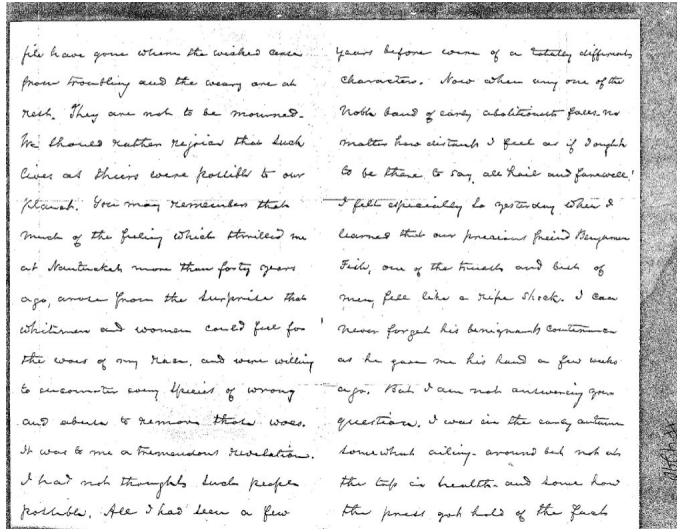
Outline of the talk

- Captatio benevolentiae
- Some fundamentals of ATR/HTR
- “Good AI”: reflecting on automating transcription
- How ATR has become an essential part of my work
- Communities of practices with ATR
- Newer Spaces of Experimentation with ATR

Graduating Seniors in 2026 chime in...



An example to start with



Frederick Douglass Correspondence

<https://christophersrose.com/2020/05/05/how-to-use-google-docs-to-ocr-arabic-text/>

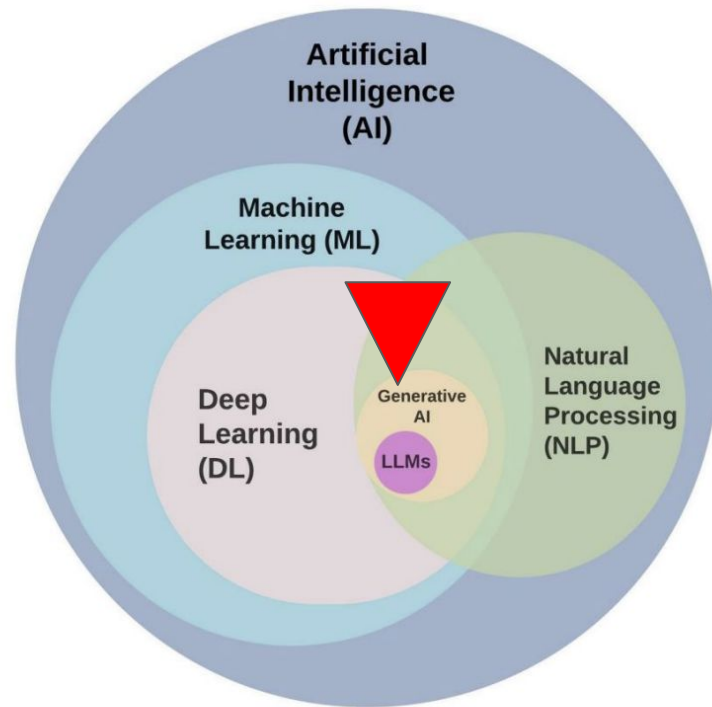


Sign at Paris restaurant

ATR in the larger picture of AI

The core models of state-of-the-art built-for-digital-scholarship platforms for ATR do not rely (now) on LLMs, but are rather based on deep learning models and transformer-based neural networks.

Post-processing can be done with LLMs. Some ATR is being done with LLMs (more on that later).



LLMs in the AI landscape, Lizama Mué,
session 2a, adapted, *red triangle is mine*

Dubreuil on AI nullifying what we believe to be humanistic practices

Q: Is there any place in the humanities where it makes sense to use AI? How might it effectively be used?

A: Like in other areas of knowledge, there are, within the humanities, some important technical elements. Take paleography, for instance, which consists of reading older systems of handwriting; or descriptive prosody, where one scans lines of regulated poetry. In such areas, and many others, self-learning automated systems and/or algorithms relying on expert knowledge could prove very useful adjuvants or fine replacements. However, one should never forget that, ultimately, the technical is theoretical: The humanities are the disciplines whose “raison d’être” is to make sense (of life, of the world, of ourselves, of sense itself). We may outsource our techniques and therefore free ourselves from tedious labor, but only if we understand that such means never were an end. ([Cornell Chronicle. 1 April 2025](#))

→ *Was transcription not truly a part of the humanities and AI has forced us to see that? Are we happy and willing to give it up transcription, or partially, to move on to more interesting tasks?*

Humanities in the Time of AI

Laurent Dubreuil



Fundamentals of ATR

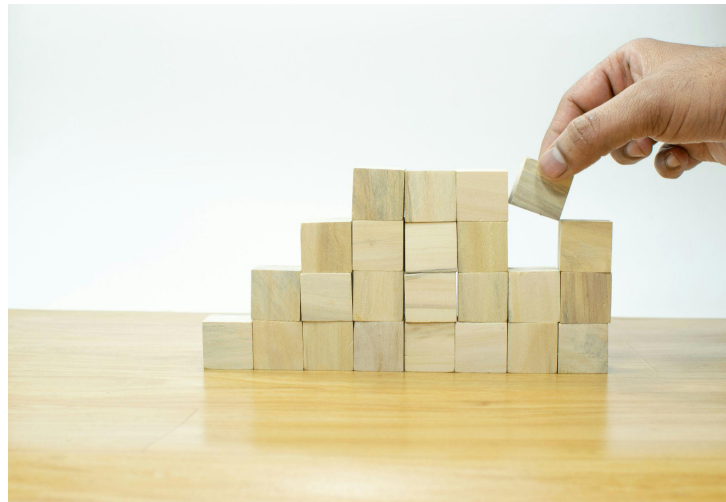


Photo by [Imagine Buddy](#) on [Unsplash](#)

3 Fundamental Steps to ATR: BAV, vat. Lat. 36

Segmentation - region detection

Polygon and baseline detection

Handwriting recognition

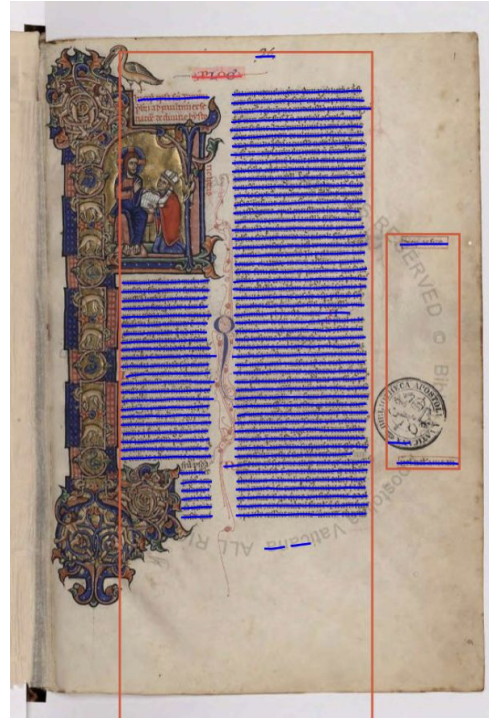


3 Fundamental Steps to ATR: BAV, vat. Lat. 36

Segmentation - region detection



Polygon and baseline detection



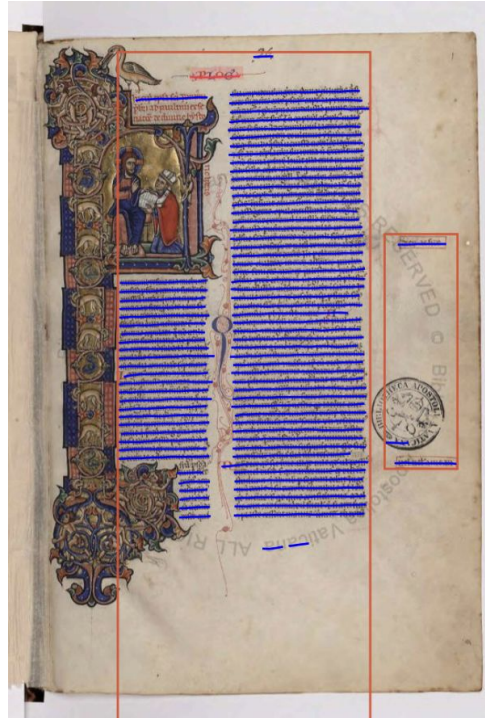
Handwriting recognition

3 Fundamental Steps to ATR: BAV, vat. Lat. 36

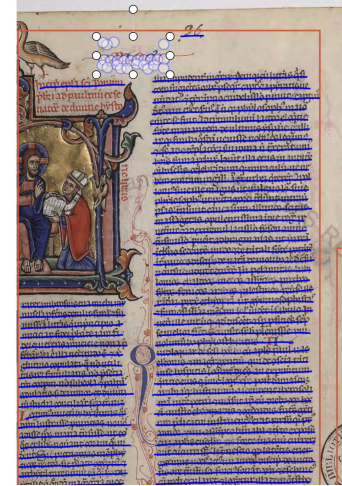
Segmentation - region detection



Polygon and baseline detection

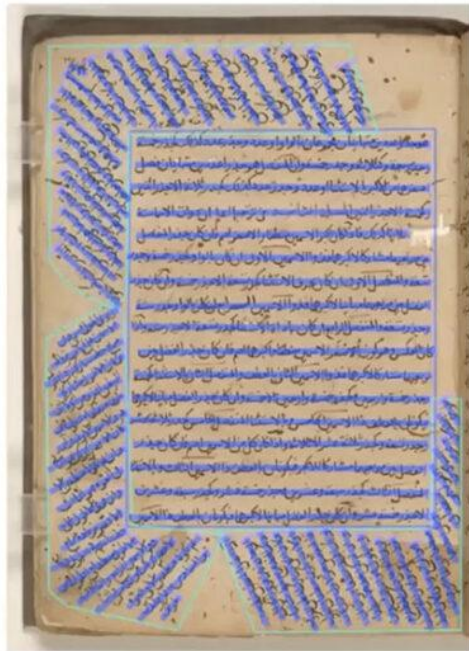


Handwriting recognition



2 S
3 fua impudent⁹ ingere. deniqꝫ cu licitā qđ ſi-
4 ncip epla fci ionimi
5 toto fugietis oꝛbe pfeqr capt⁹ 7 a pyꝛatis ue-
6 nubat 7 tyꝛāno crubellioꝫ paruit; captiū
7 oꝛt uin 7 et fu⁹ 39 qđ phylopho⁹ maio
8 emte fe fuit. Ad tytūmlibiū lacteo eloq̃tie
9 fote marientem. de ultimis ypaie galli-
10 aꝛqꝫ finib⁹ qđā ueniffē nobiles lūni
11 7 qđ aꝛ qđ tēpaletoꝫ fui roma ſi tēxat uni
12 hois fama pduū. Huut illa etus in aubity
13 oib⁹ fclis. celebrāduꝫqꝫ miraculū. ut ur-
14 bem tante inq̃ffī a ex urbez qꝛerēt apol
15 lom⁹ fua ille magus ut uulgu lo. fuie
16 phylopho⁹ ut pictagoꝛei fdunt. intraū
17 pfas; tñfunt caufa cum. albanos. fctas.
18 mafagetas. opulentiffima ī be reg pe-
19 neūit 7 aō extremū laciffio fūyon amne
20 tñiffioffo; pueit abdragmarlas. ut yareā
21 tñno fedente. aureo 7 de tātali foꝛe potatē
22 tñ paucos bicifpē de nat⁹ a de moib⁹ ac o fū

Imagine complicated layouts...



Or this one

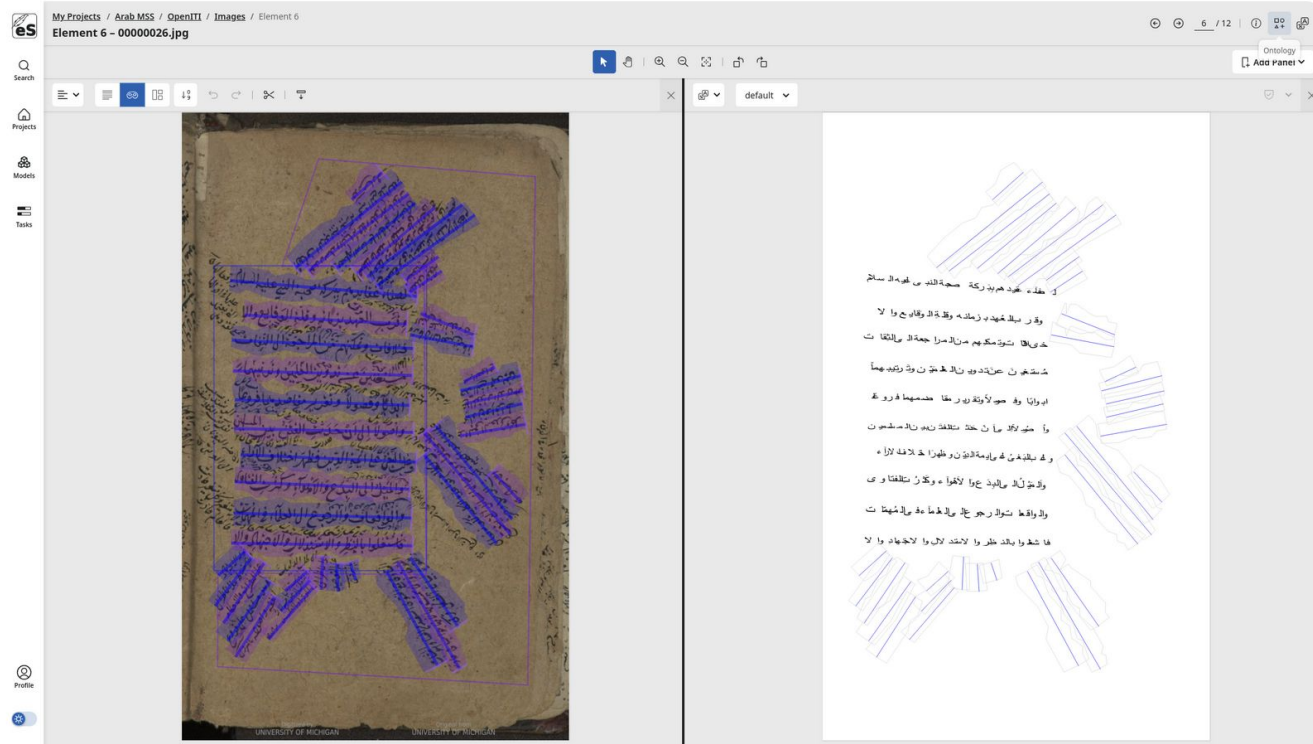


Image: e-Scriptorium

ATR as a “good” AI for use in GLAM



Photo by [Zach M](#) on [Unsplash](#)

Neudecker, Digital Curation and AI, *AI in Museums*: 2

“In order to support more advanced and diverse use of digitized collections, the digitized objects need to be processed further so as to analyse, expose, and enrich the content contained within the scans. For example, text recognition (in other words, optical character recognition [OCR] or handwritten text recognition) extracts the text from a scan to make it machine-readable, layout analysis can structure the various types of content on a scanned page into different sections like text, images, tables, et cetera, and methods from the domain of natural language processing (NLP) can be utilized to extract information (for instance, named entities) from the text or to enrich it semantically (for example, with links to a knowledge base) ... **AI methods and models have provided significant improvements for all of the above applications, for instance, for recognizing text in historical prints or handwritten documents, for document layout analysis, for content-based retrieval, or in the area of named entity recognition and linking.**”

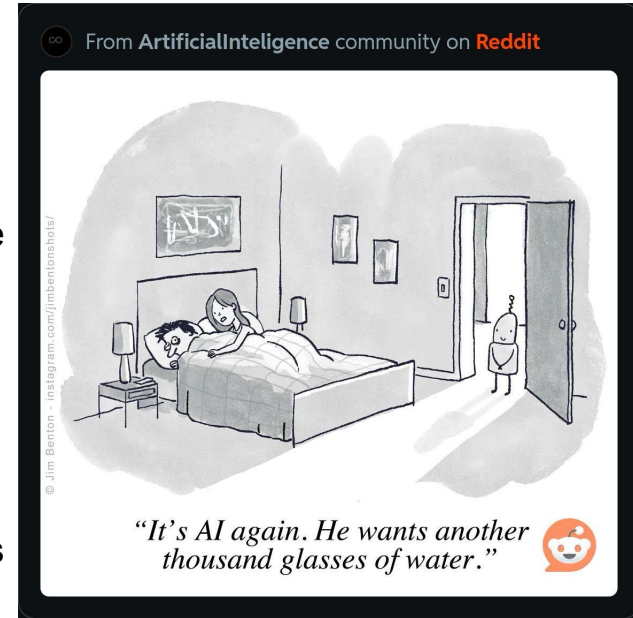
Points for approaching AI critically

As a question of knowledge, access, and representation

- How does AI produce, organize, and classify knowledge?
- **What authority should be granted to algorithmic outputs**, and how do opacity and interpretability shape human judgment?
- Whose data, experiences, and perspectives are represented, and whose are excluded?
- How do training datasets shape what can be known, predicted, and represented?
- **How does AI mediate access to information, expertise, and learning?**

As a social, political, and economic phenomenon

- What new forms of capital and platform power emerge, and who controls AI infrastructure? Where does the data go?
- **Is research in the age of AI an open field?**
- What forms of labor and power relations are embedded in AI systems?
- **What governance and regulatory structures shape AI within research institutions?**



And of the environment of course ("ArtificialIntelligence" subreddit)

Projects for which
ATR has become
an essential (even
constitutive) part of
my work



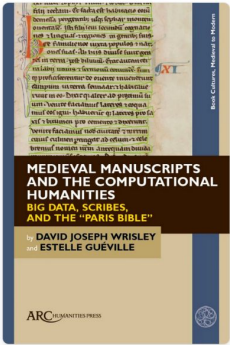
Olimpia / Unsplash

1.Paris Bible Project



Collaborative work with Estelle Guéville, Yale U

Medieval Manuscripts and the Computational Humanities, ARC Press, 2026



License

CC BY-NC-ND 4.0

Peer Review Information

This work has been peer reviewed.

[Peer Review Policy](#)

Language

Medieval Manuscripts and the Computational Humanities

Big Data, Scribes, and the “Paris Bible”

Published by [ARC Humanities Press](#)

David Joseph Wrisley , Estelle Guéville 

[Read PDF](#) [Download PDF](#) [Buy Hardback](#) [Cite](#) [Share](#)

This book examines the transformations in medieval studies—and the humanities more broadly—enabled by decades of digitization and advances in computational methods. Centring on the Paris Bible, a widely copied thirteenth- and fourteenth-century manuscript genre, we demonstrate how automated transcription produces scribal data at a scale once inaccessible, and how automation can support new approaches to localizing, dating, and contextualizing manuscripts. We argue that bringing machine learning and artificial intelligence to medieval studies not only requires re-centring expert human intelligence within computational systems, but also raises the question of the infrastructures needed for equitable, collaborative scholarship across the field. The book models how medieval studies might rethink interpretation, highlighting both the promise and risks of computational methods in manuscript research.

Metrics

Click on the tabs below to view various metrics for this book.

1,427 Abstract Views **196** Downloads **238** Reads





REUSING ATR MODELS AND FINE-TUNING THEM FOR OUR RESEARCH

Model	Transcription
Tobias Hodel model	e rex dauid senuerat habebatg letatis plit umos dies Cumaz opt ririturu ftibz no ca letiebat • D veerunt ergo ei fern sii • Queramus dno noo rege adolescentulam iurgine • e stet coram rege • et foueat eu • dor miatqz in sinu suo et calefa ciat dum nrm regem • Quelt erunt ergo adolescentulam speciosam • i omib fuubz isrl
LAD 1.0	Croc □auī□ scnuerat habcbatq ctatis plu- imos □ics Cumq□ opí ructur u- ftib□ nō ca- sciēbāt. Dfggtunt cigo q sip fui. Qucramus □ñō npo rigi a□olefcentulam uuginē; 7 ftcī co□am rege; ct foucat eū. □o□ miatq□ in finū suo ct calcta cat □nm npm rogem. Ouesi crunt ergo a□olqfcentulam spocofam. ī onmib□ fiaib□ ifrl
LAD 1.3	Cux □auī□ sciucrat habebatq□ letatis plu- unos □ies Cumq□ opi riretur□ u- ftib□ nō ca- leliebat. Oixefunt ergo ci feciu- fui. Queramus □ñō nrō regi- a□olefcentulam uirginē; 7 stet co□am rege; et foueat eū. □o□ miatq□ in sum suo et calefa- ciat □ñm nr̄m regem. Quesi erunt ergo a□olefcentulam speciosam. ī oñib□ finib□ ifrl

Proven benefits

- Managed to fine-tune ATR models in the first 2 cases to model our research questions
- Carry out scaled computational analysis directly from ATR outputs
- Able to formulate new hypotheses about manuscripts for which there is a colophon, revising older scholarship
- Able to take advantage of years of mass digitization in the GLAM sector
- Able to use many more sources than those in the open digitized record

Potential disadvantages

- fine-tuning can create the impression that research questions are being faithfully modeled, but the process may also encode new hidden assumptions
- Not working with a shared corpus, but rather a corpus modeled for our questions
- Cannot ultimately be sure of these conclusions from uncertain metadata
- Institutional collections are different from scholarly inquiry
- Cannot share ground truth and the images for full reproducibility

2. Editing Medieval Poetry with ATR



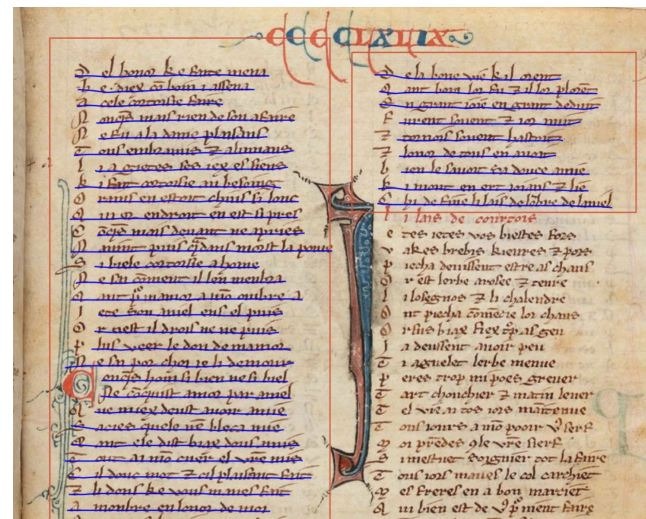
Collaborative work with Hartley R
Miller, Georgetown Qatar, HBKU Law
School

Paris, BnF ms fr 1553, “Le lai de l'ombre et de l'aniel”

ms

Digitized ms (Gallica)

Ms layout analysis (Transkribus)



Proven benefits

- Created an innovative pedagogical edition
- To pay more attention to scribal behavior.
To correct more rather than to transcribe.
HTR has exposed (Dubreuil-style) some of the assumed practices of the past
- Machine has challenged us to think critically with different workflows and new kinds of evidence.
- To make better editions because there is more time for expert judgement.

Potential disadvantages

- Could have done this by hand, but at what human cost?
- There is not consensus on what other approaches to take
- Critical reflection is often strongest during periods of novelty -> challenge of preserving critical engagement once the technology becomes normalized
- Challenging to reader for their novelty and to the community

3. Computational Study of Gulf Archives



Collaborative work with Suphan Kirmizialtin, NYU
Abu Dhabi

“Exploring Manumission Documents with Word Vectors,” JIDR 2(2024)



BRILL

JOURNAL OF DIGITAL ISLAMICATE RESEARCH

2 (2024) 1–29

Journal
of Digital
Islamicate
Research

brill.com/jdir

Exploring Gulf Manumission Documents with Word Vectors

Suphan Kirmizialtin | ORCID: 0000-0001-5020-0578

Division of Arts and Humanities, New York University Abu Dhabi,

United Arab Emirates

suphan@nyu.edu

David Joseph Wrisley | ORCID: 0000-0002-0355-1487

Division of Arts and Humanities, New York University Abu Dhabi,

United Arab Emirates

Corresponding author

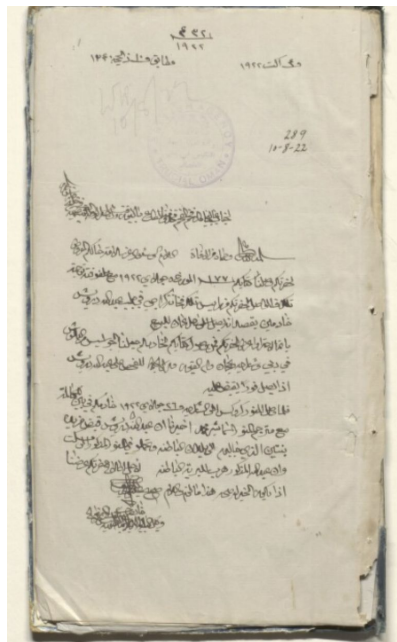
djw12@nyu.edu

Received 7 October 2024 | Accepted 26 November 2024 |

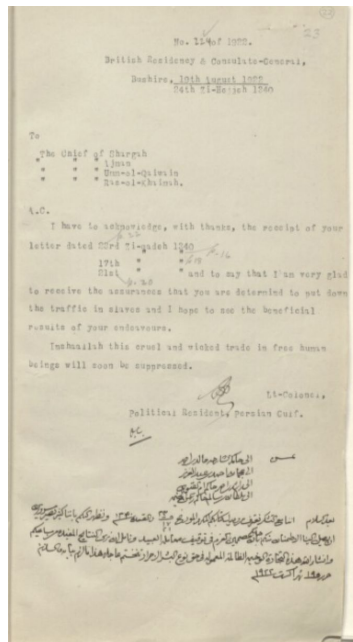
Published online 27 December 2024



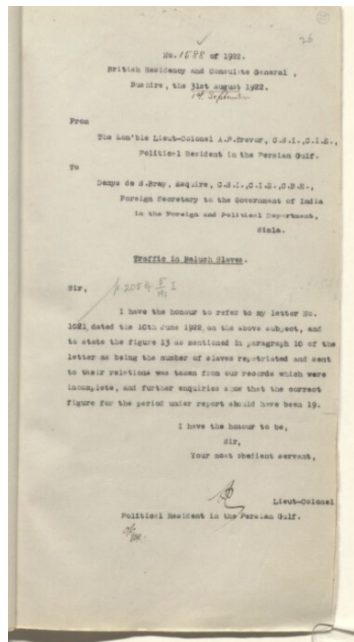
India Office Records (QDL)



File 5/191: 15



File 5/191: 56



File 5/191: 62

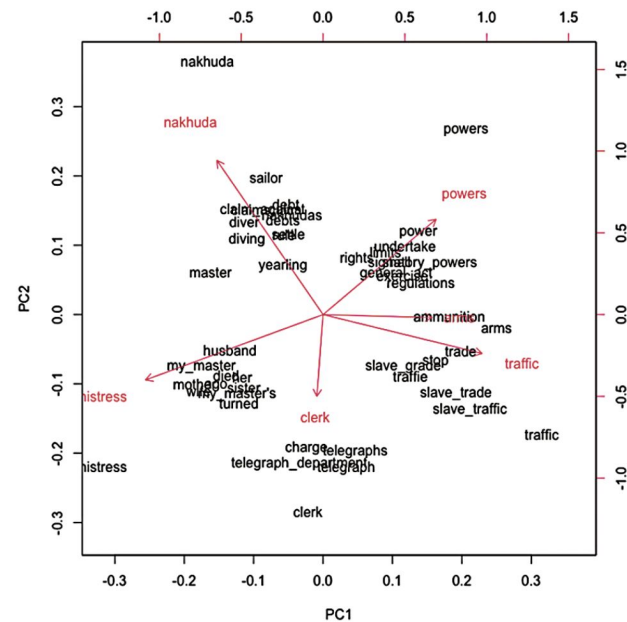


FIGURE 2 Key terms ("nakhuda" [upper left], "powers" [upper right], "weapons" [center right], "traffic" [center right], "clerk" [bottom center], and "mistress" [lower left]) in red and their contextual associations in black plotted in a two-dimensional space, generated using PCA on our custom trained wvm. The figure illustrates how specific words are spatially distributed based on their contextual similarities. The model has been trained for 2-grams: 150 dimensions, 20 iterations, 6 word window with negative sampling of 5.

Proven benefits

- Benefit from existing work of others, sometimes thousands and thousands of hours of research time, adding our own additions.
- Succeeded in working across multiple hands, as well as with typewritten English
- Powerful readings across the repetitive corpus
- Unique insights into the period and the entanglement of colonial perspectives and local experience

Potential disadvantages

- scholarly labor often becomes invisible once embedded in training data, ground truth sets, and workflows. How do we trace and cite such labor?
- Success across multiple hands and scripts often depends on substantial normalization and exclusion of other languages
- Research time cost involved in the corpus building and preparation
- Very limited and colonially-mediated corpus of digitized Arabic sources like these on which to train models

Communities of Practices

Communities of Practice

EScriptorium



Image: Wikipedia /
Chagué

Transkribus



Image: READCOOP

And there are others...

TrOCR

OLM

Calfa

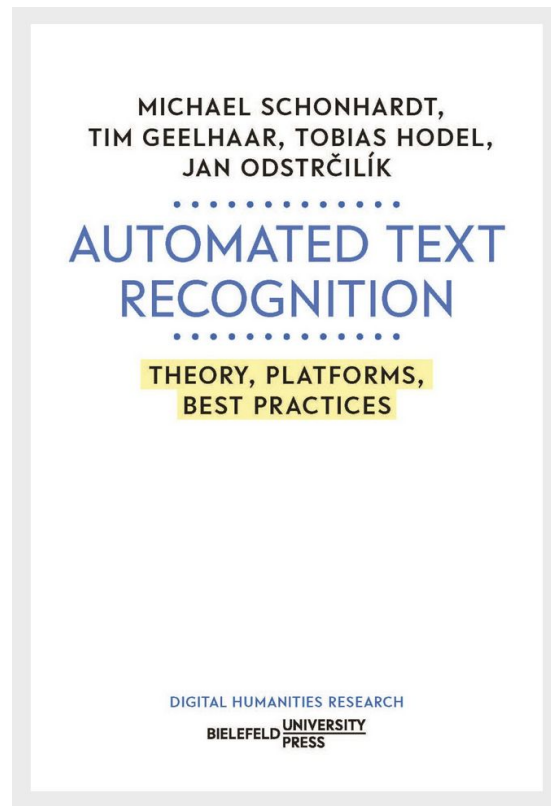
Newer Spaces of Experimentation

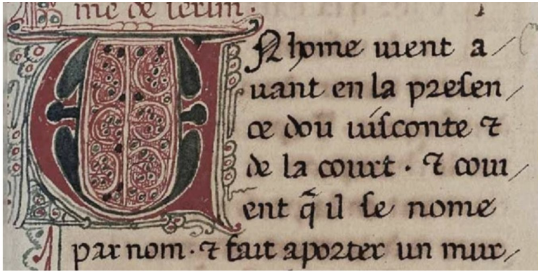


Photo by [Alex](#) on [Unsplash](#)

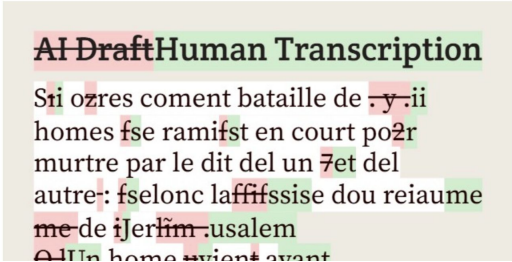
Schonhardt et al, *ATR: Theories, Platforms, Best Practices* (2026): 46

“Increasingly, transformer-based Large Language Models (LLMs) are being applied to ATR. While LLMs were traditionally restricted to post-correction, multimodal LLMs (mLLMs) now integrate vision capabilities and can directly deal with images. First approaches show promising results for well represented writing styles (English and German after 1800), though their use for complex historical corpora remains to be studied.”





Last week, I was corresponding with my friend [Laura Morreale](#) about the results of Gemini transcription on this Crusader-era manuscript. The differences between the AI results and Laura's are fairly interesting: there are some mis-readings, like *y* for *h* or *h* for *h* about half of the differences are reasonable editorial decisions. Gemini represents long s as the character *fand* retains the abbreviation of "Jerusalem" as *ierlm* while Laura modernizes both.



FromThePage: Gemini with French manuscripts

Text creation

Gemini 3 Solves Handwriting Recognition and it's a Bitter Lesson

Testing shows that Gemini 3 has effectively solved handwriting on English texts, one of the oldest problems in AI, achieving expert human levels of performance.



MARK HUMPHRIES
NOV 25, 2025



54

5

Share

JD
81,7

Benchmarking large language models for handwritten text recognition

Giorgia Crosilla
Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy
Lukas Klic
Digital Humanities Lab, Villa I Tatti, Florence, Italy, and
Giovanni Colavizza
Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy and
Department of Communication, University of Copenhagen, Copenhagen, Denmark

Abstract

Purpose – The aim of this work is to provide an overview of the current capabilities of Multimodal Large Language Models (MLLMs) for Handwritten Text Recognition (HTR), assessing their potential when compared to traditional task-specific, supervised models.

Design/methodology/approach – The approach is that of using a set of openly-available benchmarks to compare different LLMs with strong task-specific supervised baselines for the task of HTR.

Findings – The results show that LLMs currently show a strong performance on English texts, yet they demonstrate a weaker performance on languages other than English, and do not possess a significant capability for self-correction. Moreover, their comparison with Transkribus's models highlight the fact that proprietary LLM models are the best performing, in particular on modern handwriting, while for historical documents the overall performance comparison between LLMs and Transkribus is not consistent.

Originality/value – The authors are not aware of a similar study relying on open benchmarks.

Keywords Large language models, Handwritten text recognition, Automatic text recognition

Paper type Research article

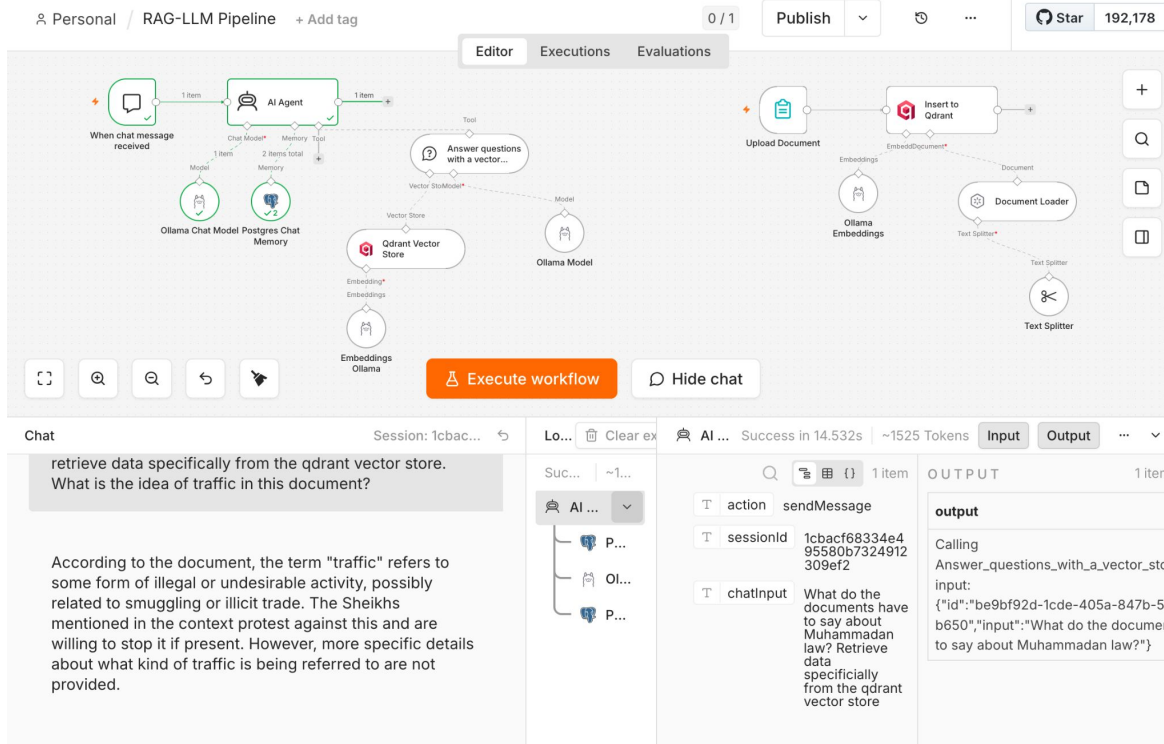
1. Introduction

In recent years, advancements in Generative AI models have begun to influence research in the Digital Humanities. The capabilities of Large Language Models (LLMs) manifest both in their ease of use for scholars, and in the wide range of applications that Generative AI can support within knowledge organization, such as the ability to classify, process and extract meaningful information from cultural data, support learning and teaching activities, facilitate collaborative



Mark Humphries, Generative History [blog]

RAG system with local LLM for querying sensitive HTR created data



Analysis of QDL
IOR/R/15/1/2
24 [“File 5/191 IV
Individual slavery
cases”](#) (1929-34)

n8n with
qwen2.5:7b

Thanks to my
DHSI class 2
weeks ago!

Final Thoughts

Expert systems leveraging AI/ML for transcription are mature systems, they are also effervescent spaces where all kinds of languages and historical scenarios are being tried out.

Hold in them the possibility of

- an enormous amount of text generation, by interested humans and by non-humans
- “potential to reshape the conditions under which scholarship is produced, for better or for worse”
- LLMs are catching up with expert systems, not only for post-processing and data processing, but for text creation → brings with it a range of problems typical of them: extrapolation, hallucination

Ethical issues

- Energy, CO₂, H₂O
- Training and deploying ATR models demands technical expertise, compute, and financial investment
- Risk of new divides in the ability to handle uncommon languages, historical state of languages (pre-1800), and scripts
- Deepening the divide between computationally privileged and underprivileged traditions and new forms of data and algorithmic colonialism (Adams)
- Increasing problems with data sovereignty
- “worsening epistemic injustice where knowledge from minority groups is systematically excluded from the discourse due to structural and technological limitations.” (Schonhardt et al, citing Melanson)
- MG -> small linguistic communities, relatively undigitized archival cultures, communities of the global South

Week 2 course: From Handwritten Document to Computable Object

| Course Offerings DHSI 2026 Week 2 (June 15–19)

From Handwritten Document to Computable Object: Critical Approaches to Automatic Text Recognition for the Digital Humanities (DHSI 2026)

🕒 30 September 2025



Description

This course will provide hands-on experience in the automatic transcription of handwritten historical documents from a variety of periods and in different languages. We adopt a critical approach to existing technologies (Transkribus, eScriptorium, and LLM-based workflows) and introduce a diversity of workflows. Using examples from Canadian and international digital collections, participants will explore how transcription and layout models can be created from scratch, as well as how existing ones can be retrained using data from the community. We also examine the importance of transcription choices on research, including when we work with models that do not directly reflect our own considerations. At the end of the course, participants will be able to assess the viability of including automatic transcription methods in individual and group research projects and they will have gained an appreciation for how automatically transcribed documents are applicable to a variety of forms of computational research. The course is intended for all levels, including those with little prior knowledge of digital scholarship. We encourage participants to bring their own research interests to the course.

اهلا وسهلا

Stop by or talk to
us!