



Critical AI Integration for Cataloging Marginalized Middle Eastern Histories

Mahdi Ganjavi (PhD)

Metadata Production and Electronic Access Specialist, University
of Toronto Libraries

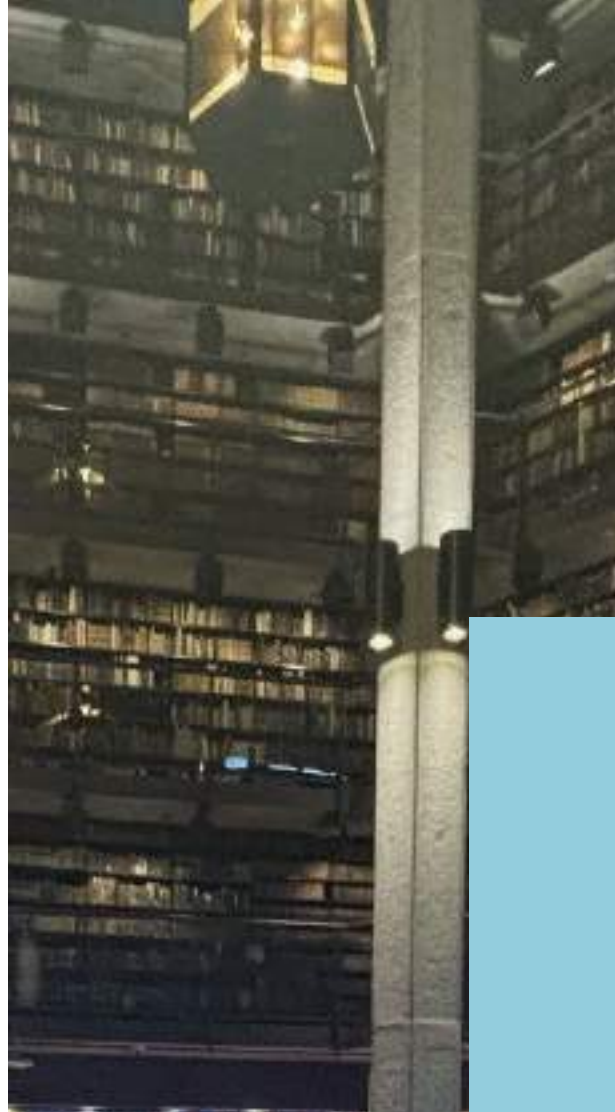
Lecturer, Faculty of Information, University of Toronto

Lecturer, Faculty of Education, University of Alberta

Institutional Context

- The Thomas Fisher Rare Book Library at the University of Toronto houses a remarkable collection of Middle Eastern and Islamic materials, reflecting the rich cultural and historical heritage of the Islamic world.
- Fisher Rare Books works closely with the Near and Middle Eastern collection at Robarts, one of the most comprehensive collections in Middle Eastern studies in North America. The collection primarily focuses on Arabic, Persian, and Turkish, with expanding holdings in Urdu and Kurdish. It also includes works in Syriac, Aramaic, and Assyrian.

Photo from the *Global Protests Archived* display at the Fisher Rare Book Library, featuring Kurdish and Persian rare materials (April 15, 2024).



Counter Archiving Activism

- Rise of personal and community archives as forms of counter-archiving activism.
- Internet platforms used to preserve Kurdish culture and history outside state control.



Shahrzad Mojab's collection on Iran's student activism

Shahrzad Mojab's collection of transnational Middle Eastern student and grass root activism (1950s-1990s): a case of modern endangered collection, which documents pivotal moments and movements in the recent history of the Middle East.

Photo: On the class struggle in China, a Persian translation of a text by the Revolutionary Communist Party, USA.



Shahrzad Mojab's Collection on Iranian Transnational Student Activism: The Student Movement and Archival Suppression in Iran

- Since the formation of the first university in Iran in the 1930s, students have been a persistent opposition force to the state.
 - Their demands and politics have evolved over time, taking a radical turn in the early 1970s.
 - The history of this resistance movement has not been adequately documented due to state repression.
 - Under both the Pahlavi monarchy (1925–1979) and the Islamic theocracy (1979–present), activists were persecuted for possessing print materials, photographs, voice recordings, or banned books.
 - Any archival efforts were seen as acts of political resistance.
 - The Islamic regime has selectively archived the student movement, emphasizing Islamic tendencies, a process we call the "Islamization of collective memory."
-



Shahrzad Mojab's Collection on Iranian Transnational Student Activism: State Control, Archival Manipulation, and Resistance

- The Iranian state, whether monarchical or Islamic, has consistently employed suppression or patronage to control the student movement.
 - Under the monarchy, pro-Shah student groups were politically and financially rewarded, while under the Islamic regime, Islamist student groups receive state support and archival preservation.
 - The state uses digital archival techniques to monitor and control radical social movements, including those of students, socialists, feminists, and workers.
 - The Islamic regime's archival efforts serve three purposes:
 - Erasing secular, socialist, nationalist, and democratic influences from history.
 - Portraying non-Islamic movements as foreign and non-authentic.
 - Establishing Islam as the sole legitimate force against injustice.
 - Given this suppression, the archival work of figures like Shahrzad Mojab in preserving the student movement's history is invaluable.
-

Flow chart of process



Templates

Fields mapped to CONSER RDA Metadata profile
One template per collection/language
Incorporates instructions of structured/unstructured data



Inventory created in Excel

Detailed inventory can be created by a non-cataloguer or generated from purchase list



If in non-Roman script – Transliterate using ALA Romanization tables

Assess use of tools like AI to support transliteration



MARC records transformation

Use AI to organize and standardize metadata
Load records to Alma
Assessment and clean-up of metadata using MARCEdit

Experimenting with ChatGPT for Transliteration of Persian Titles

Mahdi Ganjavi 10/7 1:39 PM Edited



Hi all. I thought to share an experiemnt i have done with chat gpt. In my recent tests using ChatGPT for transliteration, explored both converting transliterated text into the original Persian and using original Persian to produce transliterate based on the Library of Congress Romanization standards. Below is a summary of the results from both directions.

Conversion from Transliteration to Original Persian:

I provided the transliteration "ittihādīyah-yi kumūnist-hā-yi īrān," and ChatGPT correctly converted it into Persian as:
اتحادیه‌ی کمونیست‌های ایران

This is a highly accurate representation of the original text, suggesting that ChatGPT handles this direction well.

Other Examples:

I tested several other texts with similarly successful results:

1. **Transliteration:** "mabāni-yi qānūni-yi sarkāb-i zanān dar jumhūrī-yi islāmi-yi īrān"
 - **Original Persian:** مبانی قانونی سرکوب زنان در جمهوری اسلامی ایران
2. **Transliteration:** "muqaddamah-i barā-yi shinākht-i chihrah va jāyghāh-i zanān va mubārizāt-i ānha dar īrān"
 - **Original Persian:** مقدمه‌ای برای شناخت چهره و جایگاه زنان و مبارزات آن‌ها در ایران



Why AI-driven Transliteration is Sometimes Mistaken (case of Arabic and Persian)

1. Shared Script Between Persian and Arabic

- Both use Arabic script; models mix them.
- AI sees overlap in multilingual data.

2. Specific vs. General Romanization Systems

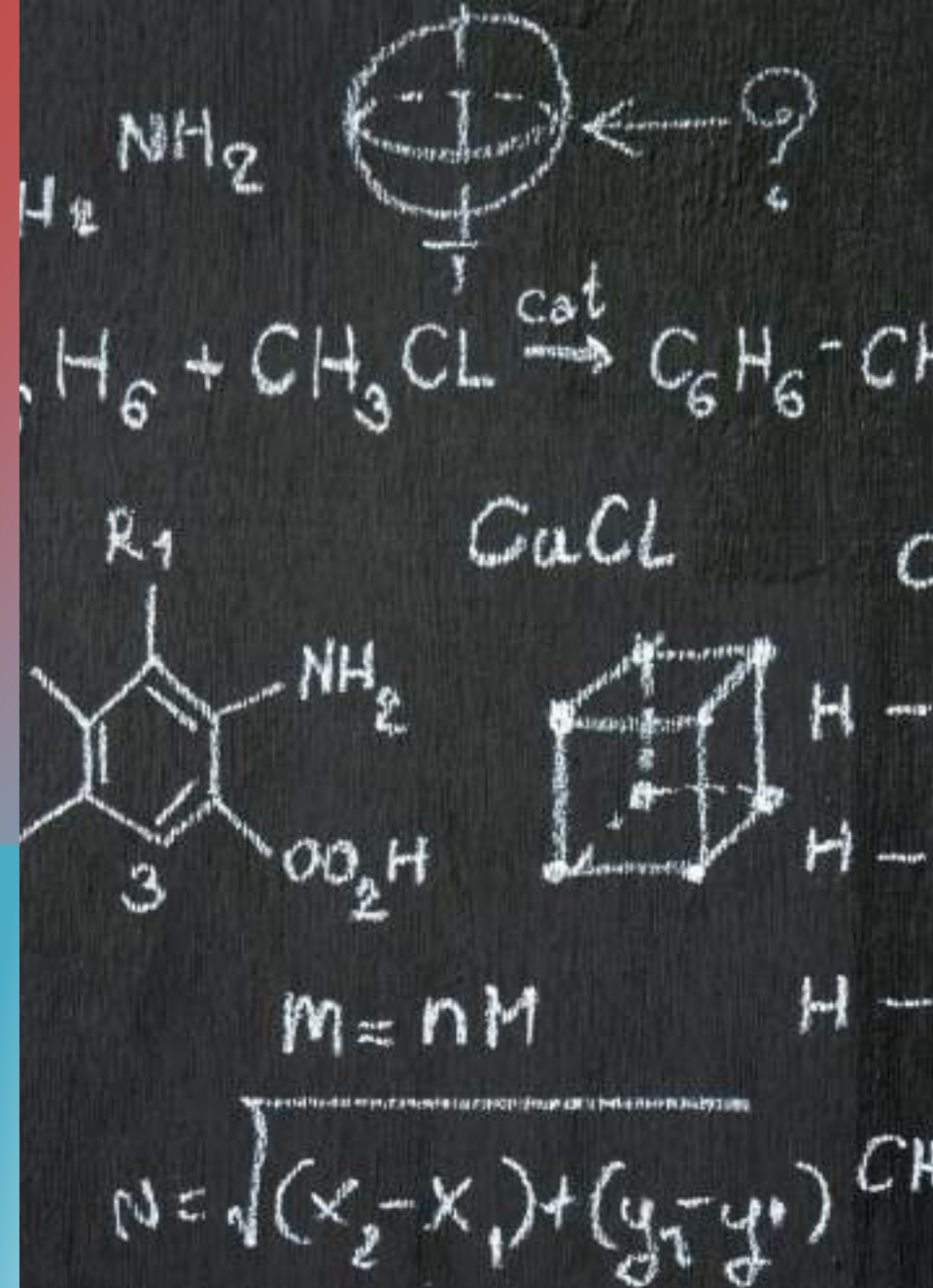
- LC Persian differs from Arabic rules.
- AI learns from mixed/informal systems too.

3. Prompting and Model Bias

- Without clear LC Persian request, AI blends styles.
- Defaults to common or popular conventions.

4. Contextual Tilt Toward Arabic

- Arabic loanwords tilt AI toward Arabic romanization.
-



Possible Mistakes-Need for Human Expert Oversight

1. Short Vowels

- ُ = o (wrong), should be u.
- ِ = e (wrong), should be i.

2. Arabic Rules Mixed

- dh for ذ (wrong), should be z.
- th for ث (wrong), should be s.

3. Confusing ض and ظ

- ض = z, ظ = ʒ.

4. Missing Macrons

- a, i, u instead of ā, ī, ū.

5. و as 'w'

- Should be 'v' in Persian.



Possible Mistakes-Need for Human Expert Oversight (2)

6. Skipping Ezafe

- Missing -i linking.

7. Alef Errors

- ‘arz (wrong), should be arz.

8. Dropping ‘ for ع

- Missing ayn marker.

9. Treating Arabic Loanwords as Persian

- Misromanizing ال as al- instead of āl-.

10. Missing Gemination

- mudat (wrong), should be muddat.

Student Journals & Grassroot Pamphlets: Iran's transnational Student and Oppositional Movement (1950s-1990s)



Corporate body/Organization (Transliterated)	Corporate body/Organization (Original language)	Title (transliterated)	Title (Original language)	Subtitle (transliterated)	Subtitle (Original language)	Variant title (Transliterated)	Variant title (Original language)	Statement of responsibility relating to the title (Original language)
Kunfidirāsyūn-i Jahān-i Muḥaṣṣalīn va Dānīshjūyān-i Īrānī (Ittīhādīyah-i Millī), SSeauthor.		nāmāh-yi pārsī		urgān-i taḥqīqī-yi kunfidirāsīyūn-i jahānī				Confederation of Iranian Students (National Union)

کنفدراسیون جهانی

نامه‌ی پارسی

ارگان تحقیقاتی

What This Case Study Suggests?

- AI must operate under human authority: In multilingual and politically vulnerable collections, automation can assist—but only within clearly defined, reviewable, and reversible constraints.
- Metadata is not neutral infrastructure: Cataloguing encodes power, language, and historical legitimacy; uncritical automation risks reproducing the very erasures counter-archives exist to resist.
- Critical AI can expand access without distortion: When embedded in template-driven workflows and guided by domain expertise, AI can reduce clerical burden while preserving linguistic, political, and archival specificity.

Questions and Discussion



THANK YOU!



CONTACT:
MAHDI.GANJAVI@UTORONTO.CA