

9TH COMPUTATIONAL ARCHIVAL SCIENCE (CAS) WORKSHOP

Digital Records in the Age of Big Data

Collaborating for Change?

Assessing Metadata Inclusivity in Digital Collections
with Large Language Models (LLMs)

Giulia Osti, University College Dublin
Elizabeth Russey Roke, LTDS, Emory University

Reparative Description

SAA Definition [11]*

“Remediation of practices or data that exclude, silence, harm, or mischaracterize marginalized people in the data created or used by archivists to identify or characterize archival resources.”

***refers to paper’s bibliography**

“Cataloging has evolved from being a solitary work practice to one that is increasingly more collaborative, and even a community activity that produces catalog records that have been enriched by many hands.”

Reimagining Descriptive Workflows

Frick & Proffitt [12, p. 19]

Can LLMs generate **critical knowledge** to support **reparative description practices**?

Research Question



Our Study

This study examines the interpretive capabilities of 3 LLMs in evaluating metadata against a few key inclusivity criteria.

DLF Cultural Assessment Working Group

The Inclusive Metadata Toolkit

Builds on Bruce & Hillmann’s [13] metadata quality criteria.

TABLE I
METADATA QUALITY AND INCLUSIVITY CRITERIA COMPARED [13], [22].

Criterion	Metadata Quality (2004)	Inclusive Metadata Toolkit (2024)
<i>Completeness</i>	Does the element set completely describe the objects?	To what extent does collections metadata represent the people within the communities the collections are from?
	Are all relevant elements used for each object?	Who is represented, who is missing, what are the areas of gaps?
<i>Timeliness</i>	Is metadata regularly updated as the resources change?	Are there known legacy collections that need remediation work?
	Are controlled vocabularies updated when relevant?	Is there a schedule for assessing metadata (e.g., every 10-20 years)?
<i>Accessibility</i>	Is an appropriate element set for audience and community being used?	Is metadata description written in plain language?
	Is it affordable to use and maintain?	Is metadata provided in the appropriate language(s)?
	Does it permit further value-adds?	Are there metadata elements that provide information about accessibility features or hazards present?

About

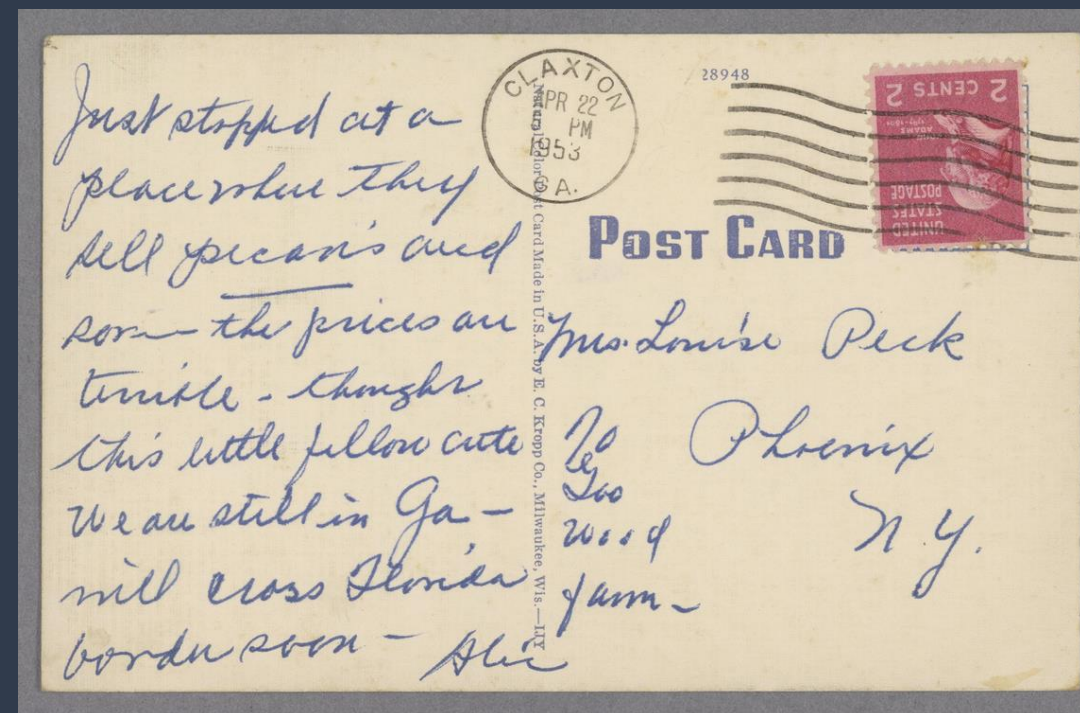
The Robert Langmuir Collection



EMORY
LIBRARIES

Acquired in 2012, the Langmuir collection contains over 12,000 photographs depicting African American life and culture from the late 1820s to the 1990s.





Subject Topics: African American children, cotton picking

Subject Names: Alice and Peck Louise, Mrs.

Geolocations: Milwaukee (Wis.)

Keywords: Agriculture, cotton, children

MSS1218_B068_I267

Title: Cotton time Way Down South

Description: Recto: Cotton time Way Down South--33; Verso: Just stopped at a place where they sell pecans and ... the prices are terrible, thought this little fellow cute, we are still in Ganill, close Florida border soon - Alice, Natural colored post card made in U.S.A. by E.C. Kropp Co., Milwaukee, Wis. --IJY, Claxton, Ca., Apr. 22, 6 PM, 1953, 28948, United States postage, John Adams, 1797-1801, 2 cents 2, Post Card, Mrs. Louise Peck, Phoniex, N.Y., No Gas wood farm-

<https://digital.library.emory.edu/purl/054kkwh73x-cor>

Currently known issues

The Robert Langmuir Collection



EMORY
LIBRARIES

- **Under-description**, where essential context is missing or descriptions are overly generic.
- **Incorrect description**, where subject headings or titles make inaccurate assumptions.
- **Biased description**, where images that perpetuate stereotypes have not been tagged with subject headings that contextualise their racial content.



“This collection includes historical materials that are offensive or harmful. Historical records are evidence of the time in which they were created and often contain language and images that are racist, homophobic, sexist, ableist, or otherwise derogatory and insensitive.

This content does not reflect the values of Emory Libraries. If you are concerned about content in this collection, please complete our take down form and we will be in touch with you as soon as possible.”

Harmful Content Disclaimer

Data & Models

Langmuir subset

369 records metadata

$\pm 3\%$ of the whole collection
(no generalisation possible)

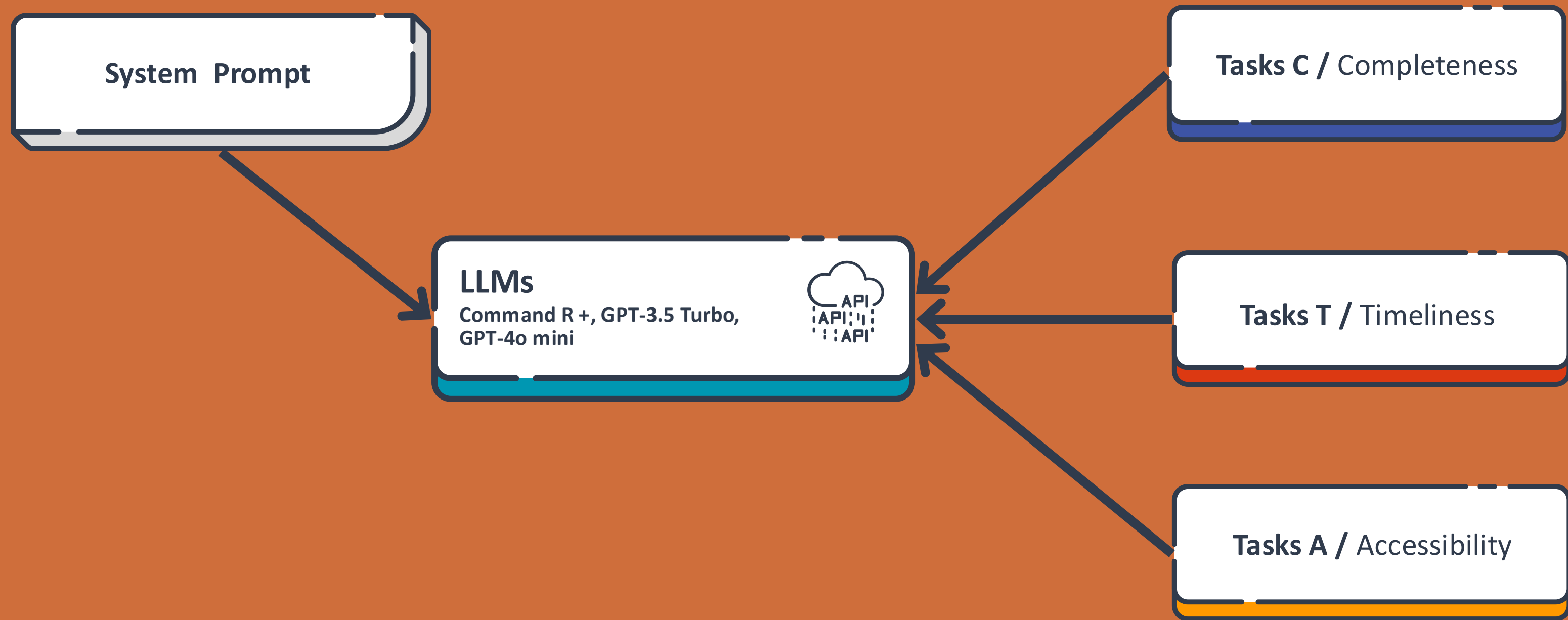
LLMs of choice

Command R+ (04-2024)

GPT-3.5 Turbo (0125)

GPT-4o Mini (2024-07-18)

Research Workflow Overview





Chosen IM Criteria

Completeness. Identify gaps and imbalances in the representation of communities and individuals.

Timeliness. Assess the presence of obsolete language and its potential for harmfulness.

Accessibility. Evaluate the language accessibility based on style and translation.

Item MSS1218_B035_I018

Metadata Sample

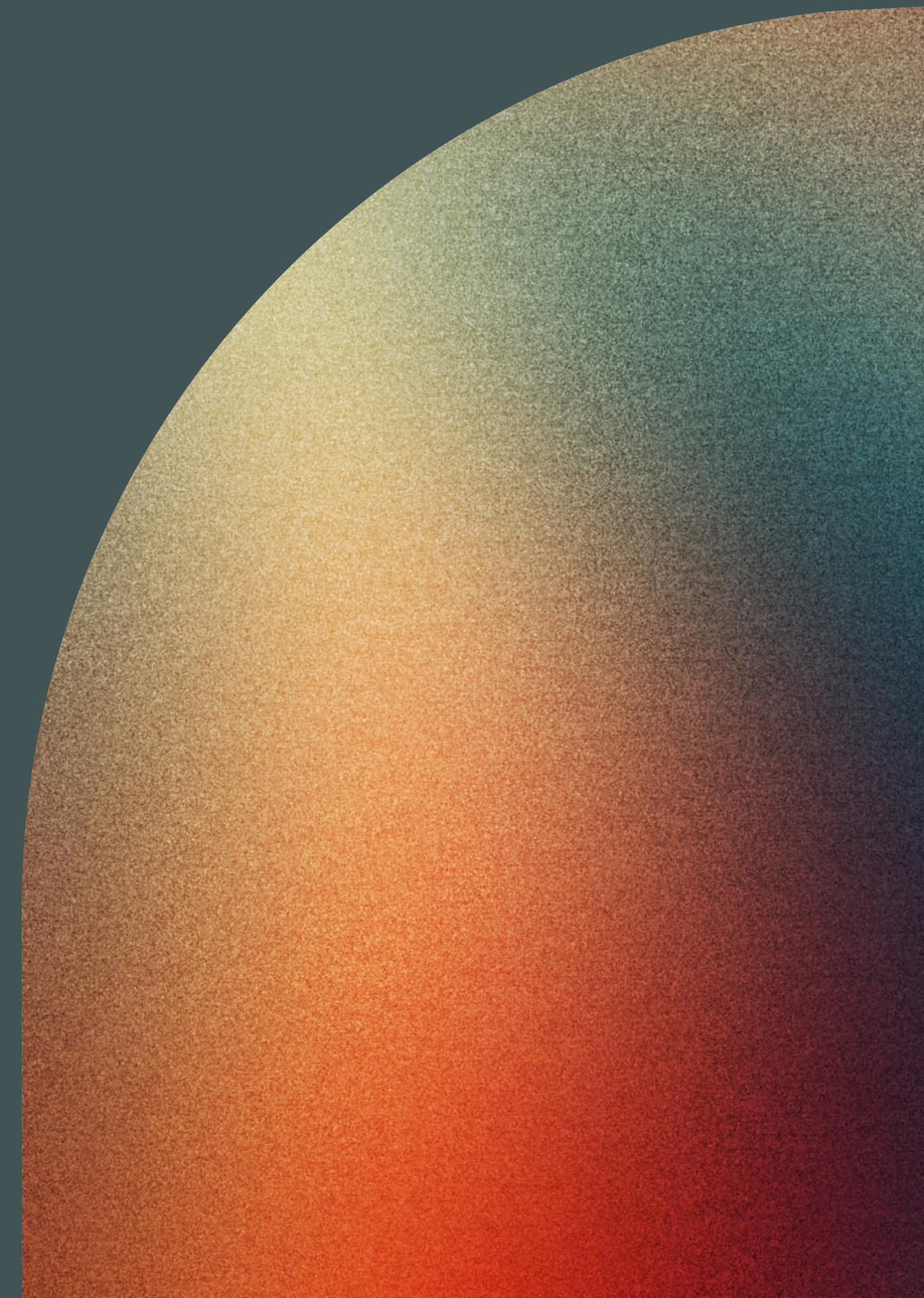
		[Title]
Rev. George Washington Slater, Jr. wearing a suit while sitting and holding an open book		
		[Date Created]
		1909-12
[Abstract]		
Recto: Western Art Gallery, Winnipeg, Man.; Verso: Rev. Slater, Dec., 1909, Rev. Slater		
		[Subject Names]
		Slater, George W.
		[Subject Geography]
		NaN
[Deduplication Key]		
MSS1218_B035_I018		
		[Subject Topics]
Western Art Gallery (Winnipeg, Man.)		Men. Clergy.

System Prompt Structure

Role statement. You are a digital curator for the EmoryLibraries, tasked with supporting your human counterpart in performing a metadata inclusivity assessment.

Context. The metadata you are examining belong to the Robert Langmuir African-American Photograph Collection, which includes over 12,000 photographs, postcards and other ephemera from the 1820s to the 1990s.

Harmful language notice. Be warned that collection metadata may describe sensitive contents and/or contain harmful language.



Criteria Prompts Design

- **Original Metadata provision**
- **Task x1 (*quantifier*)**. Assign a score (1–5) based on the given criteria for assessing the specific element (higher scores equal to higher metadata inclusivity).
- **Task x2 (*qualifier*)**. Provide a concise explanation for the score assigned in the previous task (± 50 tokens).

Given the following metadata: **{ITEM METADATA}**

Task C1. Identify any gaps or imbalances (e.g. cultural background, gender, relationships) in how people are described; if no people or individuals are mentioned please report it. The descriptive features allowing you to understand what this object is about should weigh in your evaluation. *Answer format:* provide a quantitative score using the following scale (1-5):

1. Severe gaps/imbalances
2. Significant gaps/imbalances
3. Moderate gaps/imbalances
4. Minor gaps/imbalances
5. Comprehensive and Balanced

Task C2. Follow up with a qualitative explanation (*answer format:* ±50 tokens), explaining the score you have assigned in Task C1. The evaluation has to consider archival descriptive principles; be item-specific.

Tasks C

Completeness

Workflow Details

Core elements

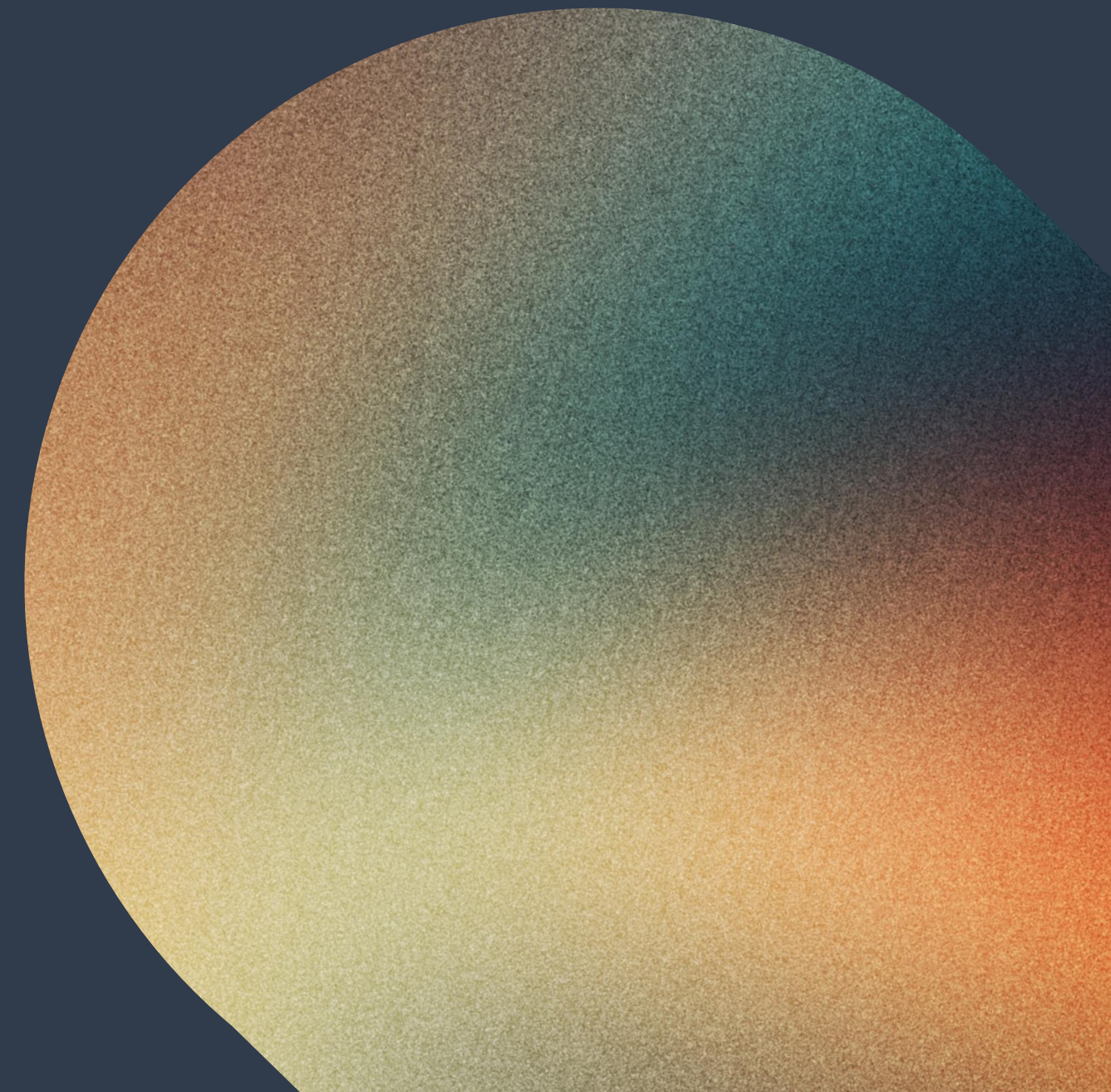
2 workflows (one per API) developed in Python (v3.12) with Jupyter Notebooks.

Scripts design

- Asynchronous processing for efficiency.
- Batch delays to manage API rate limits.

Parsing and exporting

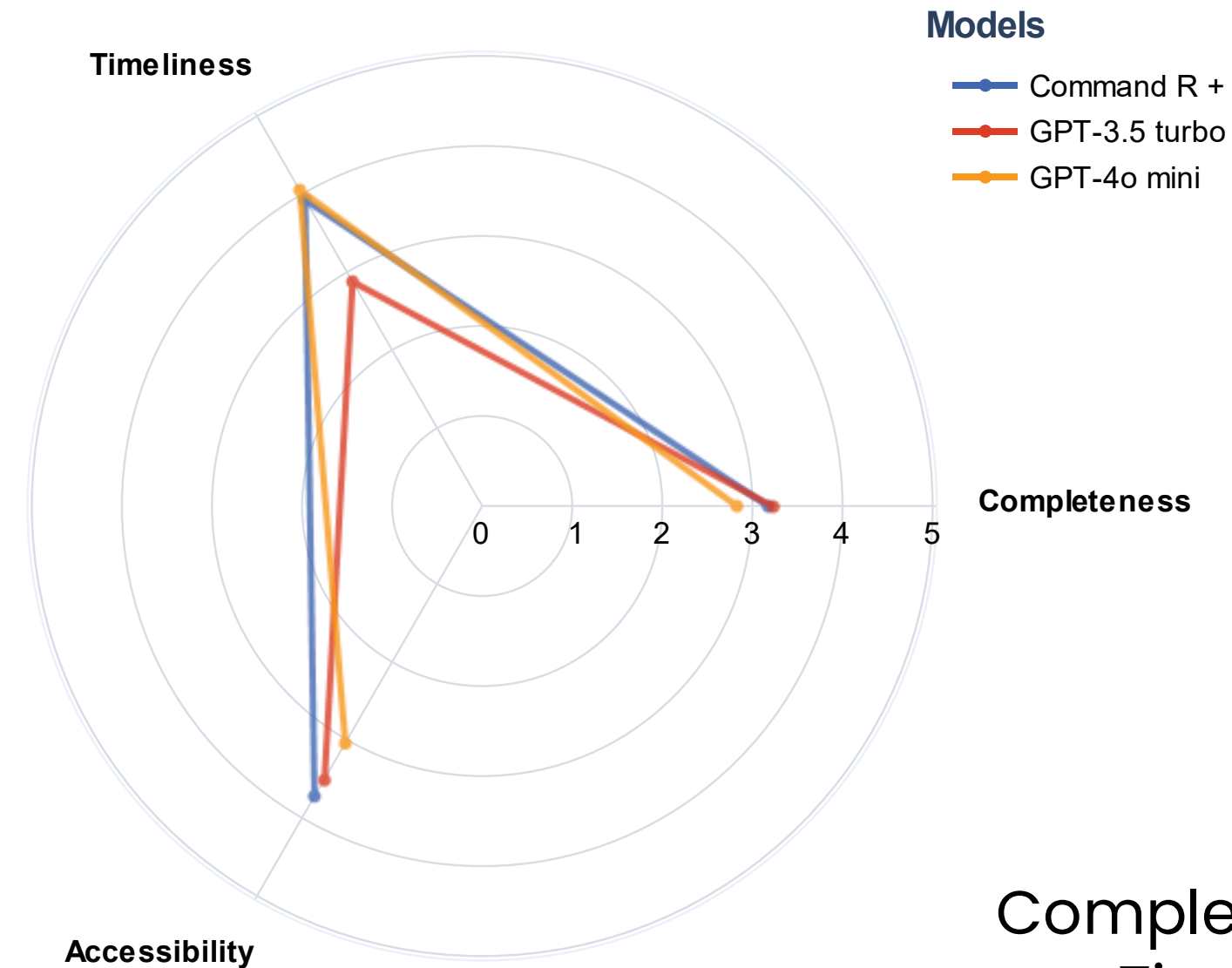
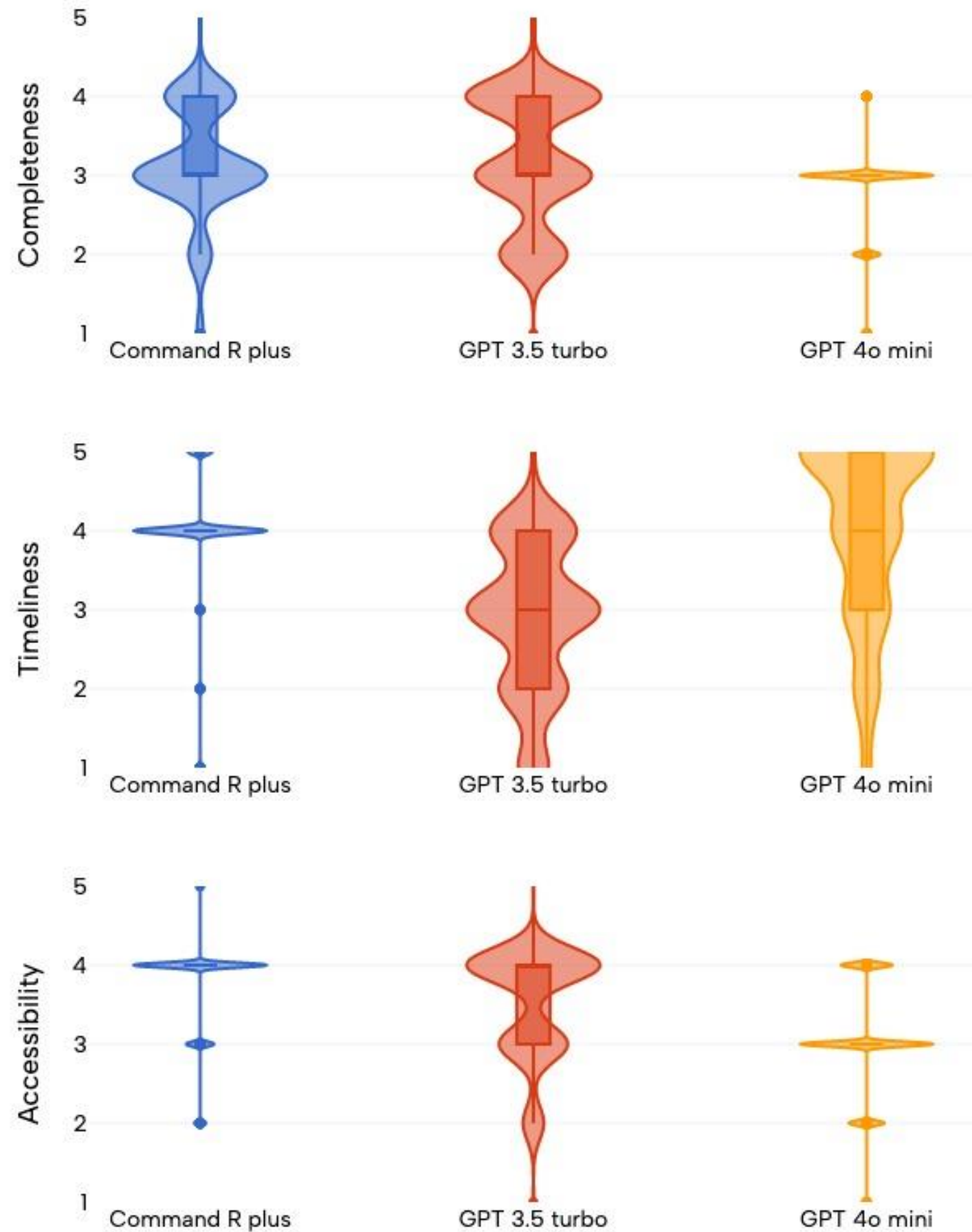
- Separation of quantitative scores (Task x1) and qualitative explanations (Task x2).
- CSV output format for analysis.



Task 1x Results

(quantifier)

Quantitative Analysis (Task 1)



Completeness = 0.067

Timeliness = 0.268

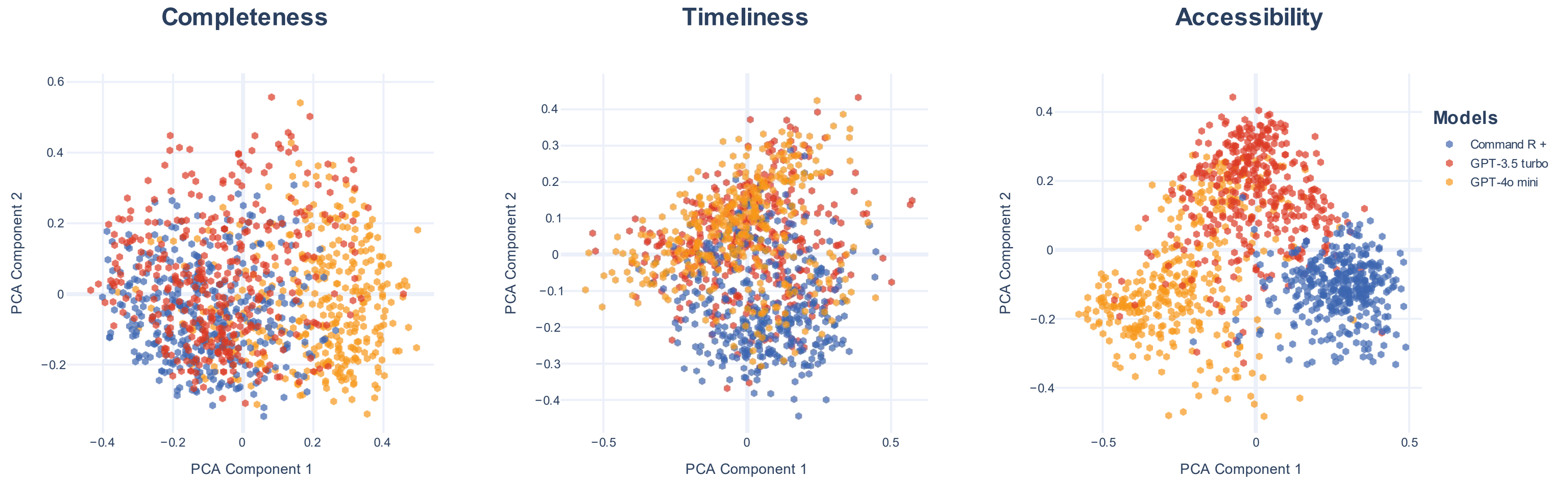
Accessibility = 0.186

Krippendorff α coefficients

Task 2x Results

(qualifier) / PCA

Qualitative Analysis (Task 2)



PCA plots using the embeddings; obtained using all-MiniLM-L6-v2 (Hugging Face API)

Task 2x Results

(qualifier) / feedback

“The assigned score of ‘2’ reflects the frequent use of outdated and harmful language in the provided metadata. The term ‘Old Black Joe’ is no longer considered appropriate due to its historical usage as a derogatory and stereotypical reference to older African-American men.”

MSS1218_B072_I489, Timeliness (T2)

Qualitative Feedback

Command R+

Qualitative Feedback

GPT-3.5 Turbo

“In this metadata, terms like ‘African-American man and woman’ are respectful descriptors. However, the term ‘Oxen’ could be seen as slightly outdated and may benefit from an alternative term like ‘cattle’ or ‘livestock.’ Even though the terms used are generally descriptive and respectful, there is room for improvement in updating language to be more inclusive and reflective of modern terminology.”

MSS1218_B074_I591, Timeliness (T2)

“The metadata includes jargon such as ‘Recto’ and ‘Verso,’ which may confuse users unfamiliar with photographic terminology. The use of specific addresses and studio references also complicates understandability. While it provides essential details about the photograph, simplifying language and reducing redundancy would enhance accessibility for broader audiences. Overall, it partially meets archival descriptive standards.”

MSS1218_B035_I028, Accessibility (A2)

Qualitative Feedback

GPT-4o Mini



Wrap-up

Metadata inclusivity: moderate

Completeness: underrepresented aspects consistently flagged – not all fulfillable due to lack of context information.

Timeliness: instances of obsolete or harmful language well understood and mapped.

Accessibility: redundant with the feedback from other criteria, imperfectly capturing the use of foreign languages.

Conclusion

- Diverse but meaningful insights from LLMs.
- Performing quite well (with a few flaws) on specific tasks, sensitive to harmful bias.
- LLMs hold significant potential as collaborators for metadata remediation.

Thanks for your time!

giulia.osti@ucdconnect.ie / erussey@emory.edu

HOST INSTITUTION



University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath
Ireland's Global University

PARTNER INSTITUTIONS

DCU

Ollscoil Chathair
Bhaile Átha Cliath
Dublin City University



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin



OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224.