

ENHANCING ACCESS TO CULTURAL ARCHIVES WITH GENERATIVE AI & KNOWLEDGE GRAPHS

RAJESH KUMAR GNANASEKARAN



Background on MSA's LoS

Why focus on Maryland State Archives' Legacy of Slavery (LoS)?

- ▶ The *Maryland State Archives (MSA)* undertook a *multi-year* effort to digitize historical records from the *LoS* project. Project to discover stories of unknown 'heroes' of slave flight and resistance.
- ▶ This effort resulted in a large, digital database containing over *400,000* records, including Certificates of Freedom (CoF), Domestic Traffic Ads (DTA), and Manumissions.
- ▶ Goal of this project: to enhance access to these digitized collections of the *LoS* datasets in an ethical and scalable manner.



Domestic Traffic Ads Dataset sample

Great Bargains.
The subscriber offers at private sale the following property, viz: two Negro Boys, one Negro Girl, about 14 years of age, two Yoke of Oxen, three Cows, thirty head of Sheep, one Still and Cap, one Whip Saw, one Gun, two Mares with fold, and one first rate Saddle Horse—All of which he will dispose of very low for cash.
SHADROCK KEENE.
feb 12 St

FOR SALE, a MULATTO MAN, 23 years of age, slave for life, and sold for no fault, only the present owner has no use for him. He is a first rate farm hand and a good miller. For farther information apply at Lewis F. Scotti's, Intelligence, Agency and Collectors Office, No. 1 West Fayette st. Basement Barnum's City Hotel. ma 7


Valuable Negroes
AT PRIVATE SALE.
The subscriber will sell at private sale a Family of as Valuable Negroes as any on the Eastern Shore, they will not be sold out of the county—viz. 1 woman & her 2 children, 1 man 38 years old, 1 do 20 years, 1 do. 23 years, 1 boy 14 years, 1 do. 6 years, 2 do. 7 years, 1 woman & her 2 children one 2 and the other 5 years, 1 woman and her child 3 years, 2 girls 4 years, 1 do. 5 years, 1 do 16 years, they are all young and handsome, and will be sold to good masters very low.
JOHN DONOVAN.
Dec. 31 Sw



Manumission Record Sample

16

BE IT KNOWN, That at the Instance
of my deceased Sister, Elizabeth Bordley, late of
Annapolis, I her Devisee, and Executor of her last
Will and Testament, have manumitted, released and
set Free, and hereby do manumit, release, and for
ever set Free, her late Negro, called *Polly*, her Sister of *Judah*,
and all her Children
(who has assumed the Name of *Molly Giles*)
_____) from the State of Slavery in
which ~~they~~ have been held hitherto under the Laws of the
Country; and with Wishes that by honest Industry
~~they~~ may thrive and preserve a good Name, I have
hereto set my Hand and Seal, this first Day of
December, 1789.

Signed, sealed and delivered, *Esau Bordley* 

in Presence of us
Walter Dulany
W. M. Brown



Certificates of Freedom Dataset sample

State of Maryland Caroline County to wit, Whereas by
the last will and testament of Sillby Bell, late of Caroline County
deceased, made and executed on the fourth day of June in
the year of our Lord Eighteen hundred and sixteen, proved
and enrolled in the office of the Register of Wills for the County
and state aforesaid, on the thirteenth and thirtieth days of July in
the year of our Lord Eighteen hundred and sixteen aforesaid
he in and by the said will gave freedom to his negro woman
to be taken place at his decease which was in a few days
after the fourth of June in the year aforesaid reference being
had to the will of the said Sillby Bell will more fully
appear; And whereas application has been made to me by a
negro woman calling herself **Lot Bell** of Caroline County,
four feet eleven inches and a half high or thereabouts, rather
straight and well made, pretty dark complexion, long face
and high cheek-bone, narrow between her temples, rather flat
nose in the face, mouth and thick lips and has a remarkably
long head of her own, has a very remarkable scar on her
head on the left side thereof which resembles a Mulberry
very much, and has no other remarkable marks to me left
having, the said negro doth swear to be about thirty years
of age, she was identified to my satisfaction by the oath
of Wm. Rachel Bell of Caroline County aforesaid —

Figure: Scanned CoF document of Lot Bell issued in Caroline County in 1816



Leveraging Generative AI Solutions

ChatLoS v1 using OpenAI GPT & RAG (Retrieval-Augmented Generation)

- ▶ With ChatGPT introduced in late 2022, we experimented with a conversational chat interface (“ChatLoS”) that pulls context from the *Domestic Traffic Ads (DTA)* dataset and provides *context-aware* responses
- ▶ Demonstrates *interactive natural language* conversational capabilities
- ▶ Great for *targeted semantic search but limited in context*
- ▶ Issues arise around *context-limitation, unable to perform data aggregation analysis or cross-collection analysis*



Understanding Retrieval-Augmented Generation

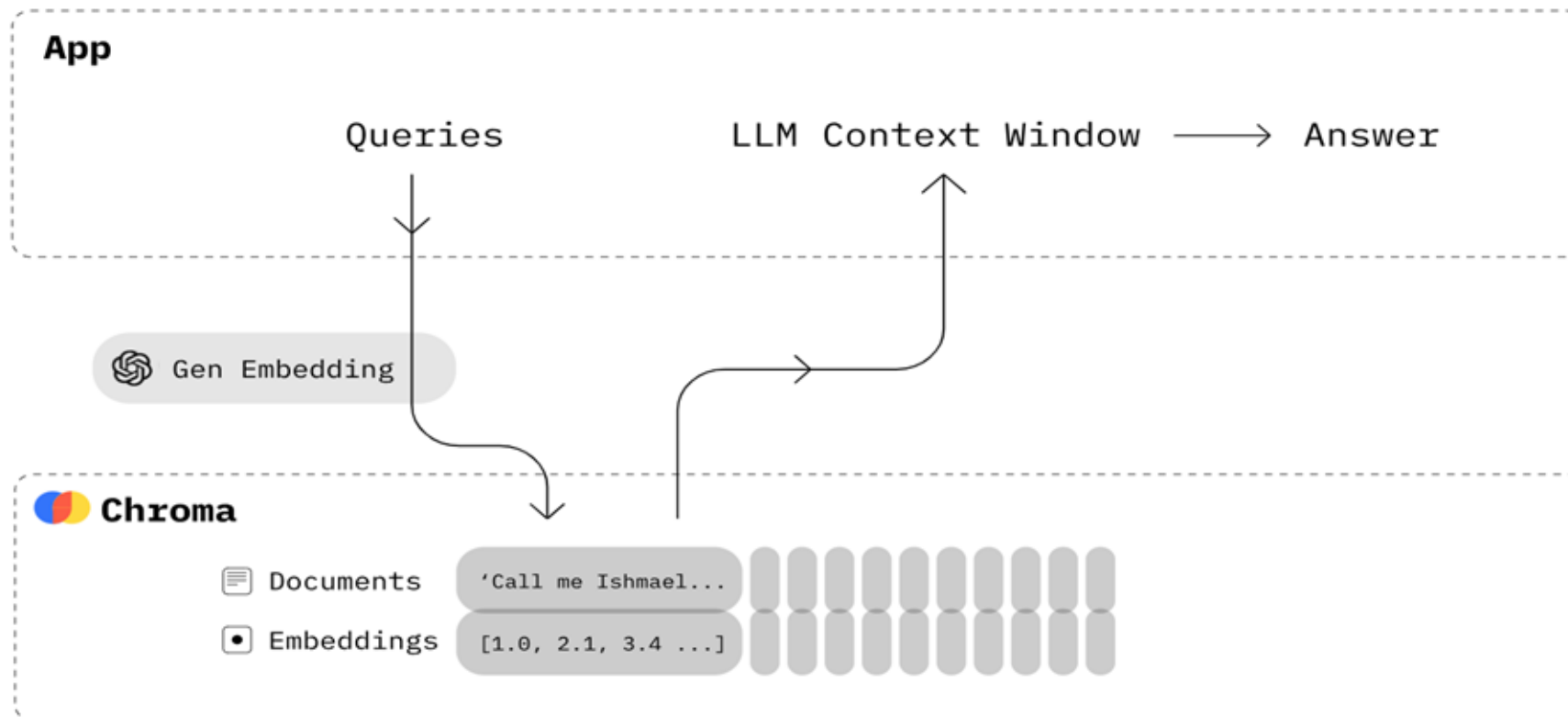
How RAG Enhances ChatLoS for LoS access

What is RAG?

- ▶ Retrieval-Augmented Generation (RAG) is a hybrid AI approach that retrieves relevant information from external sources before generating responses.
- ▶ This prevents hallucinations and ensures that LLMs provide contextually accurate answers.



Understanding Retrieval-Augmented Generation



“Explainability” in ChatLoS

ChatLoS includes an explainability layer with every response.

- ▶ "What the AI did" section clarifies query interpretation and process.
 - ▶ Human readable summary of the process
- ▶ Internal logs show source context retrieved from the dataset.
 - ▶ Internal details on how a query was performed and results were interpreted
 - ▶ Also shows the system prompt instructions added
- ▶ Promotes user trust by avoiding hallucinated or speculative answers.



Simple RAG CSV Agent

You

Is there a person named Jacob in the ads? if so list that ad

Bot

Yes, there is a person named Jacob mentioned in the ads. He appears in the advertisement published in Dorchester County on March 15, 1834. The ad offers for sale "one negro man called Jacob," along with other property, to satisfy a court judgement. The sale is noted as a public sale and was published in the Cambridge Chronicle, page 3.

Generative AI explanation behind the response

What the AI did:

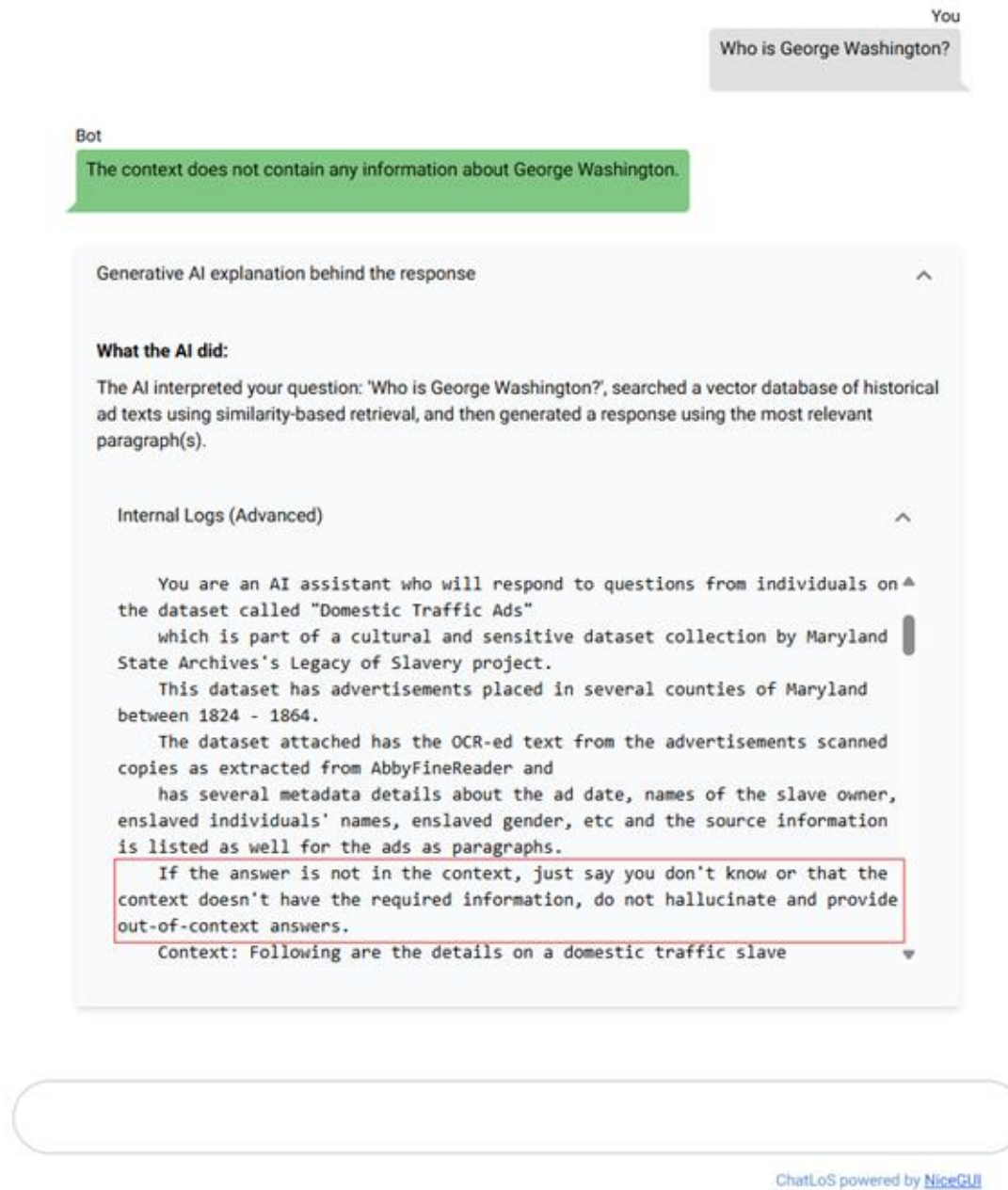
The AI interpreted your question: 'Is there a person named Jacob in the ads? if so list that ad', searched a vector database of historical ad texts using similarity-based retrieval, and then generated a response using the most relevant paragraph(s).

Internal Logs (Advanced)

Context: Following are the details on a domestic traffic slave advertisement published of the type: 'Sale' in the county - **Dorchester** on the date **March 15, 1834**. The number of people being sold through this advertisement is 1. The reason for the sale is: to satisfy court judgement and of sale disposition: public sale. The ad was published in the newspaper: Cambridge Chronicle and on page: 3. The text of the advertisement as extracted using an OCR tool from the newspaper cutting is as follows: Late Sheriff's Sale. BY virtue of a writ of venditioni exponas, issued out of Dorchester County Court, and to me directed, in the name and on behalf of Levi D. Travers, Administrator of John C. Travers, against Stewart Keene, I will sell at public sale, on Monday, the 7th day of April next, at the Court House door, in Cambridge, between the hours of 10 and 2 o'clock, for cash, all the right, title, claim, interest and demand of the said Stewart Keene, of, in and to the following property, that is to say: **one negro man called Jacob, one yoke of Steers, one Ox-Cart, one bay Horse and one gray Mare.** The

Figure: Simple RAG based ChatLoS v1 response for a sample name search question





The screenshot displays a chat window with the following elements:

- User (You):** "Who is George Washington?"
- Bot:** "The context does not contain any information about George Washington."
- Generative AI explanation behind the response:**
 - What the AI did:** The AI interpreted your question: 'Who is George Washington?', searched a vector database of historical ad texts using similarity-based retrieval, and then generated a response using the most relevant paragraph(s).
 - Internal Logs (Advanced):**
 - You are an AI assistant who will respond to questions from individuals on the dataset called "Domestic Traffic Ads" which is part of a cultural and sensitive dataset collection by Maryland State Archives's Legacy of Slavery project.
 - This dataset has advertisements placed in several counties of Maryland between 1824 - 1864.
 - The dataset attached has the OCR-ed text from the advertisements scanned copies as extracted from AbbyFineReader and has several metadata details about the ad date, names of the slave owner, enslaved individuals' names, enslaved gender, etc and the source information is listed as well for the ads as paragraphs.
 - If the answer is not in the context, just say you don't know or that the context doesn't have the required information, do not hallucinate and provide out-of-context answers.**
 - Context: Following are the details on a domestic traffic slave

At the bottom of the chat window, there is an empty input field and the text "ChatLoS powered by NiceGUI".



Leveraging Generative AI Solutions

Enhance ChatLoS using CSV Agent - ChatLoS v2

- ▶ Enhance ChatLoS by incorporating unique CSV Agent that performs a specific action
- ▶ Demonstrates interactive and complex data aggregation and advanced data analysis capabilities
- ▶ Still has some limitations - inconsistencies & inaccuracies with aggregation results with multiple datasets, unable to perform cross-collection analysis



How AI Agents Work in ChatLoS?

Enhance ChatLoS using CSV Agent - ChatLoS v2

- ▶ **User Query:** A user asks, “How many advertisements were placed in 1831?”.
- ▶ **Agent Executes Actions:** ChatLoS v2, interprets the question and determines the best approach (e.g., counting records, summarizing data). The Agent runs data analysis computations over the dataset.
- ▶ **Final AI Response & Data Visualization:** ChatLoS v2 presents the aggregated data as natural language statements with the data as structured tables.



Choose ChatLoS (Generative AI Conversational Chat Interface) version: v1 - Simple RAG, v2 - CSV Agent

Simple RAG CSV Agent

You

How many ads are there?

Bot

There are 764 ads in the dataset.

Generative AI explanation behind the response

What the AI did:

The AI understood your question: 'How many ads are there?', then scanned a structured CSV dataset, identified matching rows based on keywords or semantic relevance, and composed an answer based on the filtered content.

Internal Logs (Advanced)

> Entering new AgentExecutor chain...

Invoking: `python_repl_ast` with `{'query': 'len(df1)'}`

764There are 764 ads in the dataset.

> Finished chain.

Figure: ChatLoS v2 using CSV Agent's response for a simple data aggregate question



You: Can you create a table of the number of ads placed by each county?

Bot: The table of the number of ads placed by each county is as follows:

County	Number of Ads
Baltimore City	241
Dorchester	191
Queen Anne's	185
Baltimore	80
Allegany	41
Talbot	20
Anne Arundel	2
Caroline	2
Harford	2

Generative AI explanation behind the response

What the AI did:
The AI understood your question: 'Can you create a table of the number of ads placed by each county?', then scanned a structured CSV dataset, identified matching rows based on keywords or semantic relevance, and composed an answer based on the filtered content.

Internal Logs (Advanced)

> Entering new AgentExecutor chain...

Invoking: `python_repl_ast` with `{'query': "df1['County'].value_counts()"}`

County	
Baltimore City	241
Dorchester	191
Queen Anne's	185
Baltimore	80
Allegany	41
Talbot	20
Anne Arundel	2

ChatLoS powered by NiceGUI

You: Can you report ad counts for both public auctions and public sales?

Bot: The counts for advertisements related to sales are as follows:
- Public sale: 385
- Private sale: 374
- Public / Private sale: 5

Generative AI explanation behind the response

What the AI did:
The AI understood your question: 'Can you report ad counts for both public auctions and public sales?', then scanned a structured CSV dataset, identified matching rows based on keywords or semantic relevance, and composed an answer based on the filtered content.

Internal Logs (Advanced)

Invoking: `python_repl_ast` with `{'query': "df1['_Sale_Disposition'].value_counts()"}`

```

_Sale_Disposition
public sale          385
private sale         374
public / private sale    5
Name: count, dtype: int64
The counts for advertisements related to sales are as follows:
- Public sale: 385
- Private sale: 374
- Public / Private sale: 5
> Finished chain.

```



You

Were any ads placed on Christmas Day, if so, how many and what are the years?

Bot

Yes, ads were placed on Christmas Day. There were 5 ads placed, all on December 25, 1824.

Generative AI explanation behind the response

What the AI did:

The AI understood your question: 'Were any ads placed on Christmas Day, if so, how many and what are the years?', then scanned a structured CSV dataset, identified matching rows based on keywords or semantic relevance, and composed an answer based on the filtered content.

Internal Logs (Advanced)

Invoking: `python_repl_ast` with `{'query': "df1['_Ad_Date'].str.contains('December 25').sum()"}`

5
Invoking: `python_repl_ast` with `{'query': "df1[df1['_Ad_Date'].str.contains('December 25')]['_Ad_Date']"}`

```
13 December 25, 1824
14 December 25, 1824
15 December 25, 1824
16 December 25, 1824
17 December 25, 1824
```

Name: `_Ad_Date`, dtype: object
Yes, ads were placed on Christmas Day. There

	A	B	C	D	E	F	G	H	I	J
1	_Ad_Date	_Ad_Title	_Ad_Typ	_Age_Specific	CATEGORY --_Age_Specific	County	_Gender_Specific	_Number_of_Peop	_Press_Da	_Sale_Disposition
15	December 25, 1824	Sheriff's Sale	Sale			Dorchester	male	5	18241225	public sale
16	December 25, 1824	Sheriff's Sale	Sale			Dorchester	female	5	18241225	public sale
17	December 25, 1824	Sheriff's Sale	Sale			Dorchester	male	5	18241225	public sale
18	December 25, 1824	Sheriff's Sale	Sale			Dorchester		5	18241225	public sale
19	December 25, 1824	Sheriff's Sale	Sale			Dorchester		5	18241225	public sale
766										
767										
768										

Figure: ChatLoS v2 more examples showing broader conversational capabilities



You

How many cooks were on sale?

Bot

There were 5 individuals specified as "cook" on sale. Additionally, there were other individuals with skills that included cooking among other skills. If you are interested in the total number of individuals who had cooking as one of their skills (including those with multiple skills listed), please let me know.

You

Yes please give the counts of all the ads that were listed as cooking as one of their skills

Bot

The ads that listed cooking as one of their skills have the following counts:

- Cook, ironer, washer: 19
- Cook: 5
- Cook, washer: 2
- Cook, waiter: 2
- Cook, ironer, washer, waiter: 2
- Cook, housework: 2
- Cook, servant: 1
- Child care, cook, washer: 1

In total, there are 34 ads that mention cooking as one of their skills.

simple chat app built with NiceGUI

AG

1	NEW - Skills Specified
4	cook, servant
81	cook, housework
292	cook, waiter
293	cook, waiter
324	cook, ironer, washer
346	child care, cook, washer
517	cook
519	cook, ironer, washer
525	cook, ironer, washer
530	cook, ironer, washer
537	cook
550	cook, ironer, washer
552	cook, washer
553	cook, washer
564	cook, housework
588	cook, ironer, washer
596	cook, ironer, washer
598	cook, ironer, washer
613	cook, ironer, washer
622	cook, ironer, washer
624	cook, ironer, washer
630	cook, ironer, washer
639	cook, ironer, washer
645	cook, ironer, washer, waiter
649	cook, ironer, washer
651	cook, ironer, washer

curated-dataset-redux-OR

Ready 34 of 764 records found Accessibility: Unavailable

Filter menu options:

- Sort A to Z
- Sort Z to A
- Sort by Color
- Sheet View
- Clear Filter From "NEW - Skills Spec..."
- Filter by Color
- Text Filters
- cook
- (Select All Search Results)
- Add current selection to filter
- child care, cook, washer
- cook
- cook, housework
- cook, ironer, washer
- cook, ironer, washer, waiter
- cook, servant
- cook, waiter
- cook, washer

Figure: ChatLoS v2 more examples showing broader conversational capabilities



Leveraging Generative AI Solutions

Optimize ChatLoS using Knowledge Graph-RAG Framework - “ChatLoS v3”

- ▶ Prior solutions only looked at using one dataset in a context-aware manner and with non-AI ready data
- ▶ Knowledge Graphs (KGs) serve the purpose of getting the data “AI-Ready” by providing structured source of truth context (through nodes, relationships)
- ▶ Study included three datasets from LoS together (CoF, DTA, Manumissions) for cross-dataset insights.



How KG-RAG works with ChatLoS?

- ▶ *Knowledge Graphs (KGs)* provide structured source of truth context (through nodes, relationships)
- ▶ Nodes and relationships built by connecting multiple datasets together by integrating *CoF, DTA, Manumissions* for cross-dataset insights.
- ▶ KG-RAG connects the KG and LLM to create GraphQL or Cypher Queries dynamically and provide that to the GPT to generate natural language responses, similar to simple-RAG but more powerful
- ▶ Used Neo4J desktop to import datasets and create nodes and relationships



Node Label
EnslavedPerson
SlaveOwner
ManumissionRecord
CertificateOfFreedom
DomesticTrafficAd

Relationship Type	Source Node	Target Node
OWNED_ENSLAVED_PERSON	SlaveOwner	EnslavedPerson
PLACED_DOMESTIC_TRAFFIC_AD	SlaveOwner	DomesticTrafficAd
LISTED_FOR_DOMESTIC_TRAFFIC_SALE_IN	EnslavedPerson	DomesticTrafficAd
SIGNED_MANUMISSION	SlaveOwner	ManumissionRecord
MANUMITTED_THROUGH	EnslavedPerson	ManumissionRecord
WAS_GRANTED_FREEDOM_THROUGH	EnslavedPerson	CertificateOfFreedom
WHO_ENSLAVED_WAS_FREEDOM_FROM	CertificateOfFreedom	SlaveOwner

Database Information

Use database

neo4j

Node labels

- CertificateOfFreedom
- DomesticTrafficAd
- EnslavedPerson
- ManumissionRecord
- SlaveOwner

Relationship types

- LISTED_FOR_DOMESTIC_TRAFFIC_SALE_IN
- MANUMITTED_THROUGH
- OWNED_ENSLAVED_PERSON
- PLACED_DOMESTIC_TRAFFIC_AD
- SIGNED_MANUMISSION
- WAS_GRANTED_FREEDOM_THROUGH
- WHO_ENSLAVED_WAS_FREEDOM_FROM

neo4j\$

```

1 MATCH (so:SlaveOwner)-[:SIGNED_MANUMISSION]->(mr:ManumissionRecord),
2       (so)-[:PLACED_DOMESTIC_TRAFFIC_AD]->(dta:DomesticTrafficAd),
3       (ep:EnslavedPerson)-[:LISTED_FOR_DOMESTIC_TRAFFIC_SALE_IN]->(dta)
4 RETURN so, mr, dta, ep;

```

Figure: ChatLoS v3's Neo4J GRAPH DB setup with nodes (left), relationships (top), a sample graph (bottom)



ChatLoS enhanced with KG-RAG using Neo4j and Generative AI— Cross-dataset querying with DTA, Manumissions, and Certificates of Freedom

Can you trace the journey of individuals from being listed in Domestic Traffic Ads to receiving a Certificate of Freedom, including any manumission records?

Yes, based on the available data, we can trace the journey of an individual named John Howard from being listed in Domestic Traffic Ads to receiving his Certificate of Freedom.

- Domestic Traffic Ads:** John Howard was listed for sale in a Domestic Traffic Ad on August 15, 1829. You can view the original ad through the following source: [1829 Domestic Traffic Ad for John Howard](#).
- Manumission Records:** Approximately a year later, his Manumission Record shows that he was granted freedom on September 28, 1830.
- Certificate of Freedom:** Subsequently, he was officially recognized as a free individual through his Certificate of Freedom dated October 25, 1830.

This progression strongly suggests that John Howard transitioned from being enslaved, being listed for sale, gaining his freedom, and then receiving official recognition of that freedom. This kind of detailed tracing can further our understanding of the process of liberation during this period.

Generative AI explanation behind the response

What the AI did:

The AI interpreted your question: 'Can you trace the journey of individuals from being listed in Domestic Traffic Ads to receiving a Certificate of Freedom, including any manumission records?', generated a Cypher query to explore graph relationships in the Knowledge Graph, executed it against Neo4j, and composed an answer from the query results.

This progression strongly suggests that John Howard transitioned from being enslaved, being listed for sale, gaining his freedom, and then receiving official recognition of that freedom. This kind of detailed tracing can further our understanding of the process of liberation during this period.

Generative AI explanation behind the response

What the AI did:

The AI interpreted your question: 'Can you trace the journey of individuals from being listed in Domestic Traffic Ads to receiving a Certificate of Freedom, including any manumission records?', generated a Cypher query to explore graph relationships in the Knowledge Graph, executed it against Neo4j, and composed an answer from the query results.

Internal Logs (Advanced):

> Entering new GraphCypherQChain chain...

Generated Cypher:

```
cypher
MATCH (e:EnslavedPerson)-[:LISTED_FOR_DOMESTIC_TRAFFIC_SALE_IN]->(ad:DomesticTrafficAd)
(e)-[:MANUMITTED_THROUGH]->(m:ManumissionRecord),
(e)-[:WAS_GRANTED_FREEDOM_THROUGH]->(f:CertificateOfFreedom)
RETURN e.first_name AS FirstName,
e.last_name AS LastName,
ad.Ad_Image_Metadata AS AdImageLink,
ad.Sale_Date AS SaleDate,
m.DateManumitted AS ManumissionDate,
f.date_freed AS FreedomDate
ORDER BY FirstName, LastName, SaleDate
```

Full Context:

```
[{'FirstName': 'John', 'LastName': 'Howard', 'AdImageLink': 'http://msa
```

> Finished chain.

```
DEBUG: Chain Result: {'query': 'Can you trace the journey of individual
```

Clear Chat



Sheriff's Sale.
Will be sold on Saturday the 15th
day of August next, at the Court house
Door in Centreville,
Between the hours of 10 & 5 o'clock
of that day, for Cash
One Negroman, named
JOHN EDWARD.
(For a Term of yrs,) the property of
Stephen Boon, taken in Execution, by
Virtue of a Writ Fi. Fa. to me di-
rected, at the S of Wm. Reed, a-
gainst said Boon.
Ths. Aschom.
July 25th Sheriff.



Systematic User Evaluation using Activity Theory (AT)

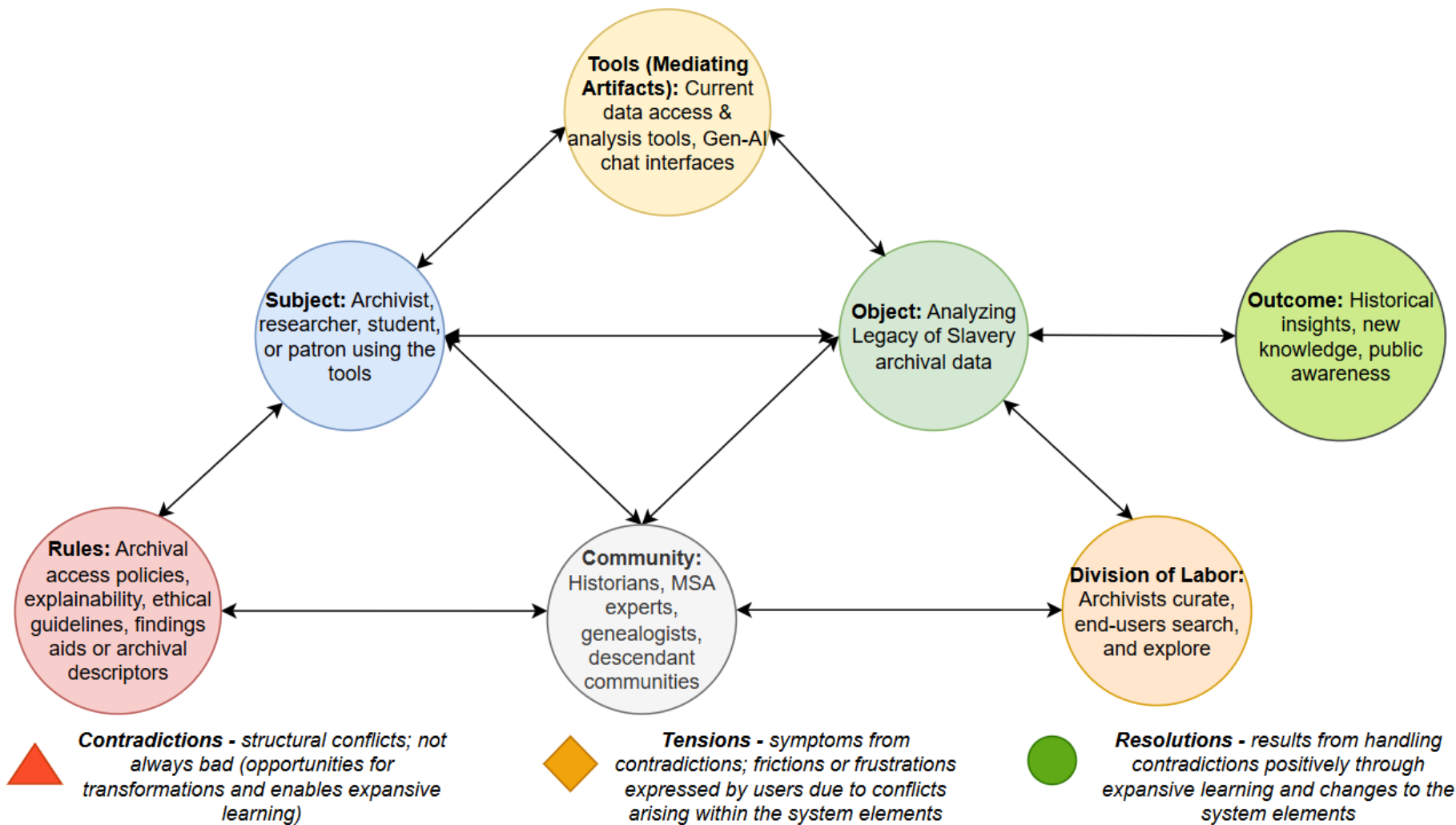


Figure: Pictorial representation of MSA's Archival Access System - An Activity Theory representation of Engestrom's Triangular model



User Study Design and Protocol

▶ Qualitative Data Collection:

- ▶ Two in-depth user interviews with MSA archival experts (identified as P1, P2)
- ▶ Semi-structured protocol covering
 - ▶ Compare current MSA access & analysis tools vs ChatLoS v1, v2, v3
- ▶ Each session lasted ~2 hours; included task trials and open-ended reflection

▶ Quantitative Data Collection:

- ▶ Comparison Rubric: 9 metrics (objective assessment)
- ▶ Trust in the Tool Survey: 7 statements
- ▶ SMEQ: Cognitive load scoring post-task (1–7 scale)



User Study Design and Protocol

▶ Qualitative Data Collection:

- ▶ Two in-depth user interviews with MSA archival experts (identified as P1, P2)
- ▶ Semi-structured protocol covering
 - ▶ Compare current MSA access & analysis tools vs ChatLoS v1, v2, v3
- ▶ Each session lasted ~2 hours; included task trials and open-ended reflection

▶ Quantitative Data Collection:

- ▶ Comparison Rubric: 9 metrics (objective assessment)
- ▶ Trust in the Tool Survey: 7 statements
- ▶ SMEQ: Cognitive load scoring post-task (1–7 scale)



Comparison Rubric: Evaluating Capabilities

Capability / Evaluation Criterion	Current MSA Data Analysis tool	ChatLoS v1	ChatLoS v2	ChatLoS v3
Natural-language question answering on a single dataset				
Interactive follow-up / conversational refinement				
Automatic semantic mapping of user terms to data columns				
Aggregate counts & multi-attribute filters				
Cross-collection linking and reasoning				
Multi-hop reasoning / trace life-paths of entities				
Explainability (surfacing code / query sent to engine)				
Typical response speed / user wait time				
Zero-coding ease-of-use for non-technical staff				

Legend: Yes = Fully supported, No = Not supported, Maybe = Partially supported

Figure: Comparison Rubric for evaluating Capabilities (nine metrics - not filled)

Trust in the Tool Survey

Trust statement	MSA tool	ChatLoS v1	ChatLoS v2	ChatLoS v3
Q1. I can rely on the system to work consistently.				
Q2. I trust the system's responses.				
Q3. The system performs accurately.				
Q4. I feel comfortable relying on the system's answers.				
Q5. The system is predictable.				
Q6. The system is dependable.				
Q7. I feel confident using the system without second-guessing it.				

Scale (1-7):

- 1 – Strongly Disagree
- 2 – Disagree
- 3 – Slightly Disagree
- 4 – Neutral
- 5 – Slightly Agree
- 6 – Agree
- 7 – Strongly Agree

Figure: Trust in the Tool Survey setup with Scale (1-7) description

SMEQ Assessing Cognitive Load

System	Score
MSA Data Access & Analysis Tool	
ChatLoS v1 (Simple RAG)	
ChatLoS v2 (CSV Agent)	
ChatLoS v3 (KG-RAG)	
Scale (1-7):	
1 = Very Low Effort	
2 = Low Effort	
3 = Mild Effort	
4 = Moderate Effort	
5 = High Effort	
6 = Very High Effort	
7 = Extremely High Effort	

Figure: SMEQ Survey setup with Scale (1-7) description

User Study Evaluation with P1: User Interview

ChatLoS - Comparative Reflections

- ▶ About ChatLoS - *“This could really make the information more accessible and engaging.”*
- ▶ **Validated core features of ChatLoS:**
 - ▶ Multi-dataset search (ChatLoS v3)
 - ▶ Natural language interface (v1–v3)
 - ▶ Source-linked, transparent outputs (v1–v3)
- ▶ **Recommended improvements:**
 - ▶ Name ambiguity handling
 - ▶ Response consistency
 - ▶ Dataset expansion (to other LoS collections)



User Study Evaluation with P2: User Interview

ChatLoS - Comparative Reflections

- ▶ About ChatLoS - *“This makes me think about how community genealogists can get more answers without our help.”*
- ▶ Emphasized different user priorities: *“accuracy, sure, but also clarity and story.”*
- ▶ Called for answer transparency: *“A good answer should give me a summary and links to original sources.”*
- ▶ Endorsed phased rollout and refinement of v3 for broader adoption

“AI is here to stay—we need to figure out how to use these tools to our advantage!”



Qualitative Data Analysis Findings from AT lens

Tool Version	Tensions	Contradictions	Resolutions	Participant(s)
MSA Tool	Literal search, no multi-filter queries; manual effort required	Subject–Tool–Object; Subject–Division of Labor	None	P1, P2
ChatLoS v1 (RAG)	Limited retrieval; no follow-up; no aggregation	Subject–Tool–Object; incomplete features; lack of continuity	Partial schema abstraction for beginners	P1, P2
ChatLoS v2 (CSV Agent)	Non-deterministic responses; occasional ambiguity	Subject–Tool–Rule; difficulty trusting inconsistent output	Accurate aggregation; semantic mapping; explanation layer	P1, P2
ChatLoS v3 (KG-RAG)	Missing fields (skills); name ambiguity; inconsistent disambiguation	Subject–Tool; Tool–Schema mismatch	Enabled life-path tracing; narrative construction; source linking	P1, P2

Figure: AT grounded qualitative data analysis findings for the two systematic user evaluation studies

Quantitative Data Collection - Results

Participant	MSA (Baseline)	ChatLoS v1	ChatLoS v2	ChatLoS v3
P1	3	3	7	8
P2	3	4	6	7

Figure: Number of rubric capabilities marked “Yes” (max = 9) for each system

Participant	MSA	ChatLoS v1	ChatLoS v2	ChatLoS v3
P1	4.14	3.00	4.71	4.71
P2	5.14	2.57	5.14	3.86

Figure 49: Mean Trust-in-the-Tool scores (7-point scale; higher = more trust).

Participant	MSA	ChatLoS v1	ChatLoS v2	ChatLoS v3
P1	5	2	2	2
P2	4	2	2	3

Figure 50: Subjective Mental-Effort Questionnaire (SMEQ) results (lower = easier)

Future Work - Community Integration

- ▶ Introduce real-time user feedback loop:
 - ▶ *thumbs up/down, comments, source flagging*
- ▶ Flipped prompting for clarification:
 - ▶ *AI prompts users when ambiguity is detected*
- ▶ Phased rollout with community-led testing: archivists, genealogists, and students



THANK YOU!



Rajesh Kumar
Gnanasekaran

