History Lab AI: A New Way to Explore a Vast Corpus of Declassified Documents - Chat Log

from Charlie Amiot to everyone: 2:17 PM
WOOOOHOOOO!

from Nick Chimicles to everyone: 2:17 PM
https://lab.history.columbia.edu/history-lab-llm

from Eleanor Mullens to everyone: 2:23 PM
The government's CUI program does still require the withholding of *some* information that, if aggregated, could pose a threat.

from Charlie Amiot to everyone: 2:23 PM
Do you think PII rules need to change, or simply how the materials are processed?

from Josh Finnell to everyone: 2:25 PM
Taking me back to my days at Los Alamos National Laboratory!

from Eleanor Mullens to everyone: 2:26 PM
CUI = Controlled Unclassified Information

from Mark Hamilton to everyone: 2:27 PM
Thank you Eleanor!

from Eleanor Mullens to everyone: 2:28 PM
UFOs, b/c they're not sure.

from Charlie Amiot to everyone: 2:28 PM
It takes longer for Bigfoot.

from Lisa Pritchard to everyone: 2:28 PM
LOL Charlie

from Eleanor Mullens to everyone: 2:28 PM
They just hold onto the nuclear records

from Charlie Amiot to everyone: 2:29 PM
@Lisa If we don't take time to laugh, we just end up crying!

from Christy McCall to everyone: 2:29 PM
How does the content compare to what's in GovInfo? Is it a different scope?

from Charlie Amiot to everyone: 2:31 PM
@Christy Not an answer, but a personal observation. I found items using this tool that I previously had no idea existed where I thought I would've.

from Eleanor Mullens to everyone: 2:31 PM
without knowing any details about the corpus, I'm assuming it's quite different from what's in GovInfo b/c the documents were released due to FOIA requests.

from Amanda Carter to everyone: 2:32 PM
LOL, filed under "Misc." That's the WORST... but also common enough in archives

from Ben Petersen to everyone: 2:34 PM
Do you have any research or data on PII detection and are you in discussions with NARA or other agencies on your findings?

from Elizabeth Williams to everyone: 2:34 PM
Speaking of collecting data, why are you requiring login for the AI interface?

from Amanda Carter to everyone: 2:37 PM
What AI product was used and how long did it take to set up and get good, accurate answers?

from Nick Chimicles to everyone: 2:37 PM
@Elizabeth Thanks for the question. We ask people to log in so we can learn how the tool is being used. That usage data helps us see what works and where it struggles.

from Amanda Carter to everyone: 2:37 PM
Do you see any hallucinations and how do you handle them if you do?

from Charlie Amiot to everyone: 2:38 PM
Tacking on to that question, what does a hallucination look like in this system, if you do get them?

from Charlie Amiot to everyone: 2:38 PM
For those newer to AI, a "hallucination" is when the AI "believes" something is true and correct, but it is not.

from Daniel Cornwall to everyone: 2:39 PM
Adding to Charlie - like making up a Department of Labor unions report with a fake OCLC number and link to a random web page.

from Nick Chimicles to everyone: 2:39 PM
https://lab.history.columbia.edu/history-lab-llm

from Eleanor Mullens to everyone: 2:40 PM
I got that same "quit" notice when I tried to download the slides, but it didn't kick me out.

from Eleanor Mullens to everyone: 2:40 PM
What is a tool call?

from Sinai Wood to everyone: 2:42 PM
not a question

from Eleanor Mullens to everyone: 2:42 PM

it wouldn't know that it's supposed to be 2 words

from Charlie Amiot to everyone: 2:43 PM
Rule #2 of AI: You get out what you put in. If you give it a trash prompt, you are likely to get a trash response.

from Trina Magi to everyone: 2:43 PM
That's funny, because sometimes at the reference desk I get two-word "questions," and I don't know what the user wants, either!

from Sinai Wood to everyone: 2:43 PM
lol

from Amanda Carter to everyone: 2:43 PM
It's helpful to give AI context to focus the request and reduce hallucinations, especially if it wasn't trained on the librarian's reference interview techniques. ;)

from Eleanor Mullens to everyone: 2:44 PM
+1 Amanda

from Amanda Carter to everyone: 2:44 PM
+1 Charlie (Rule #2)

from Charlie Amiot to everyone: 2:45 PM
The first search I ran on this site was as follows and I had a blast going through results: What evidence exists showing dishonesty on the part of Richard Nixon when putting forward to the public information about cannabis/marijuana/marihuana that was blatantly wrong?

from Amanda Carter to everyone: 2:45 PM
That sounds fun, Charlie. I think I'll borrow your question and try it!

from Susanne Caro to everyone: 2:46 PM
"I understand you're asking about the Roswell aliens. However, I am designed to provide information based on declassified historical documents. Information about extraterrestrial beings is beyond the scope of this archive."

from Bert Chapman to everyone: 2:46 PM
Being able to produce search results should have been taken care of before the presentation began!

from Eleanor Mullens to everyone: 2:46 PM
I would absolutely LOVE to create something like this (on a much smaller scale) with a trove of historical documents at my library.

from Daniel Cornwall to everyone: 2:47 PM
My prompt was a topic I had in college - Display three objections to the accession of Turkey to NATO in the early 1950s? Which countries had the strongest objections? - Reasonably happy with these casual results. Much better than stacks and stacks of paper FRUS volumes!

from Amanda Carter to everyone: 2:47 PM
You say it quotes verbatim there is no summary? That may actually help reduce hallucinations, as well. Great idea! I love that.

from Charlie Amiot to everyone: 2:48 PM
Nick, will you be publishing a paper on the programming side of this project?

from Kate Wells to everyone: 2:51 PM
How much OCR editing clean up have you needed to do due to redaction?

from Charlie Amiot to everyone: 2:52 PM
I want this database and all available CRS Reports to be combined.

from Daniel Cornwall to everyone: 2:52 PM
+1 Charlien

from Daniel Cornwall to everyone: 2:53 PM
+1 Charlie

from Lia Vella to everyone: 2:54 PM
Has the history Lab experienced any new challenges this year, given the clash between the administration and universities and Columbia in particular?

from Antoine McGrath to everyone: 2:54 PM
In the smart entity recognition process you mentioned mapping to wikidata values. Did you import wikidata information once and build on top of it, or do you utilize live wikidata? Do you weight trending topics, either user searches previously seen, or by monitoring news topics?

from Charlie Amiot to everyone: 2:54 PM
STUDENT WORKERS ARE THE LIFEBLOOD OF LIBRARIES!

from Nick Chimicles to everyone: 2:55 PM
History Lab AI: https://lab.history.columbia.edu/history-lab-llm

from Jacqui Celecia to everyone: 2:55 PM
+Charlie

from Nick Chimicles to everyone: 2:55 PM
Survey: https://docs.google.com/forms/d/e/1FAIpQLScdKqNGoGG7x-6O33bu2vkA8EUH9OHBZE4bzX89plvVPosx-g/viewform?embedded=true

from Sinai Wood to everyone: 2:55 PM
blog post url? please thank you

from Nick Chimicles to everyone: 2:56 PM
Github (history lab): https://github.com/history-lab

from Nick Chimicles to everyone: 2:56 PM
Github (History Lab AI): https://github.com/Ramus-Network/chat-agent-historylab

from Charlie Amiot to everyone: 2:56 PM
I'm just saying, UNT won digital library of the year... They have those older CRS Reports...

from Nick Chimicles to everyone: 2:56 PM
Email us! - info@history-lab.org

from Hugh Truslow to everyone: 2:57 PM
@ Sinai - I think this is it: https://newsletter.dancohen.org/archive/ai-and-libraries-archives-and-museums-loosely-coupled/

from Sinai Wood to everyone: 2:57 PM
Hugh, that is what I found but not exactly the same?

from Hugh Truslow to everyone: 2:57 PM
Not sure!

from Sinai Wood to everyone: 2:58 PM
The blog post title on the slide was Mcp future archives in the ai ecosystem

from Sarah Erekson to everyone: 2:58 PM
Is this tool ready for undergraduate history students? They are studying 1990-1998.

from Sarah Erekson to everyone: 2:59 PM
Also are they other programs that have double history computer science majors?

from Charlie Amiot to everyone: 2:59 PM
@Sarah That hurt me... 1990?!

from Amanda Carter to everyone: 3:00 PM
Thanks for those answers to my questions, very helpful!

from Nick Chimicles to everyone: 3:00 PM
Dan Cohen blog post: https://newsletter.dancohen.org/archive/ai-and-libraries-archives-and-museums-loosely-coupled/

from Sinai Wood to everyone: 3:01 PM
Okay Hugh,you were right. The slide looked different I thought. Thank you Nick too.

from Brian Garrison to everyone: 3:02 PM
Is there a sweet spot for the size of the archive? How many records does one need to gather in order for a tool like this to be useful? Or to put it differently, for some smaller states thinking about creating a history project, what's the size of the region they should be contemplating?

from Rich Gause to everyone: 3:02 PM
Limitations for 1990's coverage - which FRUS volumes have been published for those years to date?

from Raymond Hicks to everyone: 3:03 PM
Antoine, I started the NER process 6 years ago, so we use spaCy with a knowledge base we created. A large part of the issue was that the documents would not have the same personal names as Wikidata--lots of last names only or title and last name. If we started now, we would pursue a different avenue.

from Susanne Caro to everyone: 3:03 PM
I'm curious how the source material could cause the AI to be bias. Do you think it is non-bias?

from Matthew Connelly to everyone: 3:03 PM
This shows you what the National Archives/State Dept. has made available from the CFPF: https://aad.archives.gov/aad/series-description.jsp?s=4073

from Matthew Connelly to everyone: 3:05 PM
https://history.state.gov/historicaldocuments/status-of-the-series

from Charlie Amiot to everyone: 3:06 PM
We also cancelled The Public Papers of the President so...

from Hugh Truslow to everyone: 3:06 PM
Wasn't there an advisory board for FRUS that was disbanded?

from Charlie Amiot to everyone: 3:06 PM
Yes, I'm still cranky about PPP being cut.

from Angie Bonnell to everyone: 3:07 PM
+1 Hugh :(

from Amanda Carter to everyone: 3:07 PM
Is there a way to provide feedback on the responses from the AI? Like typos of inaccuracies?

from Troy Black to everyone: 3:07 PM
I loved the Public Papers of the President bound volumes.   So did our patrons.

from Lisa Pritchard to everyone: 3:07 PM
+1 Troy

from Susanne Caro to everyone: 3:07 PM
I think those State records are reviewed for sensitive material- do they have enough people to work on that?

from Lisa Pritchard to everyone: 3:08 PM
I have a little sign with the bound volumes about the fact that it is no longer printed.

from Marie Concannon to everyone: 3:08 PM
Thank you for doing this work, Matthew.

from Aimee Quinn to everyone: 3:08 PM
+1 Marie

from Amanda Carter to everyone: 3:08 PM
+1 Marie, agreed!

from Kris Abery to everyone: 3:09 PM
+1 Marie

from Angie Bonnell to everyone: 3:09 PM
+1 Marie

from Charlie Amiot to everyone: 3:09 PM
My mind was just blown.

from Nick Chimicles to everyone: 3:09 PM
https://docs.google.com/forms/d/e/1FAIpQLScdKqNGoGG7x-
6O33bu2vkA8EUH9OHBZE4bzX89plvVPosx-g/viewform?embedded=true

from Shawn King to everyone: 3:11 PM
The big 3 law database companies have incorporated AI searching and we have been attending a lot
of traininig on using their tools

from Autumn Kladder to everyone: 3:12 PM
I've been playing with it during the presentation and it's very cool, but I noticed that it only ever gives
five documents no matter the question. Even asking for more documents, it says it only has five on
that topic. Is that intentional?

from Troy Black to everyone: 3:12 PM
Matthew I think it is generational. Its also about training. What people are comfortable with.

from Charlie Amiot to everyone: 3:12 PM
Would be interesting to run a UI comparison test. Make it where a percentage of searchers must go
through a short, guided instructional tour before searching.

from Sinai Wood to everyone: 3:12 PM
+1 Troy. Early days yet.

from Barbara Moreno to everyone: 3:12 PM
Are you considering expanding the scope of type of federal documents that you are using. Like SEC
documents, etc.?

from Marie Concannon to everyone: 3:13 PM
In our Docs stacks, we have a set of books (in the Y3s I believe) listing adverse nuclear events. In
these books, I read about radioactive milk being given to boy scouts to see what would happen to
them. So many incidents like this. But try to find the equivalent to this set on the internet, good
luck! I have not been able to find anything.

from Amanda Carter to everyone: 3:13 PM
I work with researchers who want to set up something like this for state gov docs so it's definitely on the radar. I love what you've done here as it's a great example of the possibilities! <3

from Charlie Amiot to everyone: 3:13 PM
Nick, isn't that agentic AI?

from Daniel Cornwall to everyone: 3:13 PM
@Marie - EEK.

from Kristina Bobe to everyone: 3:13 PM
For older documents, how does it handle alternate place names like Porto Rico for Puerto Rico, etc.?

from Charlie Amiot to everyone: 3:14 PM
It handled all the MANY variations on cannabis. I was really impressed.

from Charlie Amiot to everyone: 3:14 PM
CRS REPORTS

from Charlie Amiot to everyone: 3:14 PM
CRS. REPORTS.

from Troy Black to everyone: 3:14 PM
Would HathiTrust be willing to work with you.

from Cass Hartnett to everyone: 3:14 PM
Very eye-opening!

from Daniel Cornwall to everyone: 3:14 PM
Thank you!

from Angie Bonnell to everyone: 3:14 PM
Thank you!!!

from Frank Lester to everyone: 3:14 PM
Thank you all so much!

from Kaylee Kern to everyone: 3:14 PM
Thank you!

from Troy Black to everyone: 3:15 PM
CRS Reports please include

from Charlie Amiot to everyone: 3:15 PM
+1 Troy. Hathi.

from Kris Abery to everyone: 3:15 PM

Thaks Matt, Ray and Nick. Excellent presention

from Aimee Quinn to everyone: 3:15 PM
Thank you so very much. Fascinating presentation.

from Rhonda Murphy to everyone: 3:15 PM
Thank you.

from Sandy Himel to everyone: 3:15 PM
Thank you.

from Charlie Amiot to everyone: 3:15 PM
Thank you!

from Josh Finnell to everyone: 3:15 PM
Great presentation!

from Perveen Rustomfram to everyone: 3:15 PM
Thank you!

from Michael Dent to everyone: 3:15 PM
Thank you!

from Min Shaheen to everyone: 3:15 PM
Thank you very much..

from Cynthia Etkin to everyone: 3:15 PM
Maatthew, Nic, Ray, excellent!

from Eleanor Mullens to everyone: 3:15 PM
thanks all.

from Hugh Truslow to everyone: 3:15 PM
Really fantastic -- thank you so much!

from Andrea Stelljes to everyone: 3:15 PM
Thank you!

from Jose Torres to everyone: 3:15 PM
THANKS!

from Charlie Amiot to everyone: 3:15 PM
Kelly is this Room A?

from Ashley Dahlen to everyone: 3:15 PM
We will resume at 3:30 Eastern

from Troy Black to everyone: 3:16 PM

How do I geto Room B

from Charlie Amiot to everyone: 3:16 PM
Go back to the iCohere landing page.

from Charlie Amiot to everyone: 3:16 PM
You'll exit this room completely.

from Troy Black to everyone: 3:17 PM
Thank you @Kelly

from Scott Matheson to everyone: 3:17 PM
https://secure.iCohere.com/FDLP-Fall2025