

ai + history collaboratory

For discussion at today's workshop

Colin Greenstreet

Tuesday, December 9th 2025

I want to structure today's session into three parts

Section A

1. Rapid fire introductions
2. Goals of this season's collaboratory
3. Working together
 - a) Synchronous
 - i. Transcripts
 - b) Asynchronous working
 - i. Discussion board
 - ii. Wiki
 - iii. Code
 - iv. Substack
 - v. Outreach
 - c) Personal workflows

Section B

1. Socratic Research Assistant
 - Goals
 - Design approach
 - Demo
 - Play

Section C

1. Homework

I am proposing three ambitious goals for this season's collaboratory – **experiment** and **build tools** and **find our voice** – and propose we pursue them over six (or possibly seven) sessions

Proposed topics

December 2025	Socratic Research Assistants
January 2026	HTR/NER
February 2026	Search/Discovery
March 2026	Databases/Agents
April 2026	Simulations
May 2026	Skills/Tools
June 2026	History hackathon [for discussion]

My plan is that we work together synchronously once a month, use our Github to work asynchronously between synchronous sessions, and that we each attempt to build learnings into our personal workflows

- Synchronous
 - i. Transcripts
- Asynchronous
 - i. Discussion board
 - ii. Wiki
 - iii. Code
 - iv. Substack
 - v. Outreach
- Personal workflows

Here is a quick look at our collaboratory GitHub resources – discussion board, wiki and code

Click [here](#) for the ai-and-history-collaboratory GitHub

Click [here](#) for the discussion board

Click [here](#) for the wiki

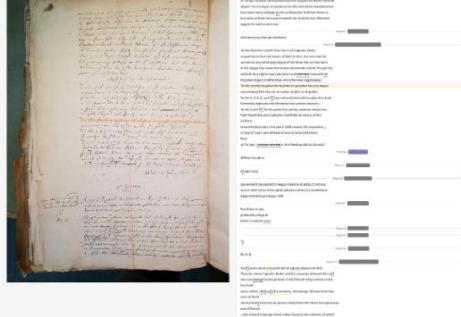
Click [here](#) for code [Socratic Research Agent]

The screenshot shows Colin Greenstreet's GitHub profile. At the top, there is a pinned repository card for "HCA" (Public) which contains deposition volumes from the English High Court of Admiralty. Below this is a pinned card for "Generative-Lives-Research-App" (Public template) which supports a Substack site. Another pinned card is for "ai-and-history-collaboratory" (Public), an online forum for historical research. The main profile area includes a photo of Colin with his dog, his name, title ("Research public historian and founder of MarineLives"), and a link to his profile. It also lists 11 followers and 40 following. Below this is a "Contributions" section showing activity for the year 2025, with a chart showing contributions per month. A "Contribution activity" section shows a single commit made on December 3, 2025, to the repository "Addaci/ai-and-history-collaboratory". The "Achievements" section shows one achievement: joining the "ai-and-history-collaboratory" organization. The "Organizations" section lists three organizations he is part of. The bottom of the page has a "Show more activity" button and a note about unexpected GitHub profile guides.

I am developing and publishing many of my ideas for the ai + history collaboratory in my Generative Lives substack

Generative Lives

[Home](#) [Notes](#) [Archive](#) [About](#)



A New Lens into the Archive

You are in an archive. You find a document in a language you don't understand. You take a photo, input it into Gemini 3 Pro. 60 seconds later you ha...

DEC 4 • COLIN GREENSTREET

What's top of my mind?

My Generative Lives agenda for the next three months

OCT 30 • COLIN GREENSTREET



Subscribed

Introducing myself: Colin Greenstreet, research public historian

A user's perspective on generative ai for historical research and archival practice

OCT 26 • COLIN GREENSTREET



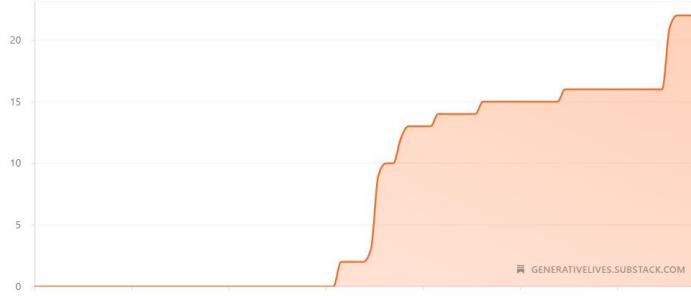
ai-collaboratory

marineLives-ai-collaboratory

Overview

90 days [...](#)

Total subscribers	90d views
22	1.9K



20
15
10
5
0

11 Sept 24 Sept 7 Oct 20 Oct 2 Nov 15 Nov 28 Nov

GENERATIVELIVES.SUBSTACK.COM

Latest post [View stats](#)

 A New Lens into the Archive
4 Dec • Colin Greenstreet [...](#)

Drafts [View all](#)

Building a research assistant to help write a G... [Draft](#) [...](#)
Edited Nov 21, 12:12 PM

Bluesky postings as an archival object [Draft](#) [...](#)
Edited Oct 28, 10:41 AM

How to use the Substack editor [Draft](#) [...](#)
Edited Oct 26, 9:39 PM

Untitled [Draft](#) [...](#)
Edited Oct 22, 1:34 PM

[Share post](#)

Recent posts [View all](#)

Post	Subscribers	Views	Opened	Actions
A New Lens into the Archive 4 Dec • Colin Greenstreet	3 Subs	1.28k Views	81% Opened	...
What's top of my mind? 30 Oct • Colin Greenstreet	1 Subs	77 Views	64% Opened	...
Introducing myself: Colin Greenstreet, research public historian 26 Oct • Colin Greenstreet	10 Subs	168 Views	50% Opened	...

I am using email and Bluesky for outreach to doctoral candidates and faculty

← Rhys Kaminski-Jones
@rhyskamjones.blsky.social ...

Yesterday at 12:42

Good afternoon Rhys. I enjoyed chatting with you publicly on Bluesky in response to Tim Hitchcock's post a few days ago. Would you be interested in exploring the topic further? I appreciate Old/Middle Welsh is a low resource language, but over the weekend I have had some very interesting results from looking at text fragments from scholastic Bengali, which themselves are derived from Sanskrit. I currently have my work being looked at by a computational linguist, and hopefully by a Bengali and/or Puranic philologist. I would be happy to share the work, and at this stage am NOT asserting as to the level of its accuracy, but the methodology is interesting (I used two instances of Gemini, each independent and unknowing of each other) and a third instance as a forensic palaeographer and philologist analysing the work of the other two.

The specific text fragment I used is from the British Library Endangered Archives scheme from the Sundarban Anchalik Sangrahashala Collection [18th-19th century] eap.bl.uk/project/EAP759, accessed 07/12/2025

12:44

For what its worth (and it may be worth nothing) in the context of your own linguistic speciality, I asked a couple of basic questions to Gemini 3 Pro Preview after asking it to adopt a relevant persona (or at least I hope it is relevant). You can read my question and the provisional answer here. You are obviously the expert, so I won't attempt to judge the content, though the exploratory process and initial hypotheses at what is clearly potentially only the very start of a conversation with Gemini struck me as interesting: gemini.google.com/share/668be3...

12:51

← Rhys Kaminski-Jones
@rhyskamjones.blsky.social ...

Today at 08:27

Thanks, it was an interesting discussion. Those specific issues re medieval Celtic I wouldn't be the best person to work with, you'd need a linguist (I work on modern reception of these texts). I would be interested in using Gemini to do a draft transcription of the English-language manuscript travel journals I'm going to start working on next year if you have advice on training/workshops etc for how to go about that though (we used Transkribus for a print transcription project I'm currently working on, but I'm getting the impression Gemini has trumped it on handwritten sources)

08:27

Re: Celtic, it's likely medieval Irish specialists who might be most interested as that's the language with large amounts as yet untranslated. But my overall worry about this technology continues to be that it will start attracting people who don't want to be careful with properly checking their results, especially when many of the departments that actually teach medieval Irish are under threat. There are multiple departments currently under threat or understaffed, so the idea of AI translation might feel understandably more like a threat than an opportunity, given how so many universities are now being run!

08:35

← Rhys Kaminski-Jones
@rhyskamjones.blsky.social ...

<notebooklm.google.com/notebook/092...> 10:29

Thanks that's very useful - getting going on this in the new year, so if you're willing might throw a few questions your way!

11:48

I would be delighted. You are also welcome to join or drop in on our ai + history collaboratory. github.com/Addaci/ai-an...

The Jan 20th 2026 session has the following agenda:

Developments in HTR and NER and their implications for historical research

Discussing Gemini 3 Pro Preview's native English language HTR and NER capabilities

Discussing Gemini 3 Pro Preview's low resource language HTR and NER capabilities.

Use case: Armenian; Mixed Russian/Armenian

Use case: Scholastic Bengali

Use case: Yiddish

Reviewing theoretical basis of strong Gemini performance in machine transcription

Sharing experimental data for transcription of English High Court of Admiralty depositions

Contrasting Gemini 3 Pro Preview results with Leo and Transkribus

Sharing experimental data for transcription of primary source handwritten documents used by other Ai + History collaboratory members

Crosilla G, Klic L, Colavizza G (2025), "Benchmarking large language models for handwritten text recognition". Journal of Documentation, Vol. 81 No. 7 pp. 334–354, doi: doi.org/10.1108/JD-0..... Click here

Greenstreet, Colin, 'A New Lens into the Archive: You are in an archive. You find a document in a language you don't understand. You take a photo, input it into Gemini 3 Pro. 60 seconds later you have a transcription, transliteration, and translation', Generative Lives, December 4th 2025. Click here

Humphries, Mark, '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.', Generative History, November 24th 2025. Click here.'



13:15

I am developing many of my ideas for the ai + history collaborative in Gemini 3 Pro Preview (Consumer Bot & AI Studio API), especially on the topics of HTR/NER [session 2], search/discovery [session 3], and databases/agents [session 4]

Yiddish Palaeography Workflow ... ⓘ
AI Infographics for Historical Do... ⓘ
Analyzing Generative History Su... ⓘ
Scholar Labs: AI for Research ⓘ
Simulating Nested Learning in Gemini
AI Agents in Historical Research
Medieval Celtic LLM Data & Experime...
AI-Assisted Historical Research Colla...
JSON Prompts for Gemini Consumer
VLM-First Approach for Historical Ma...
Ladino HTR and Translation Pilot
Forensic Analysis of Translated Text F...
Gemini's Thinking Process & Forensic...
Ottoman Turkish Document Analysis
Armenian Article Context and Publica...
Yiddish Script, Unicode, and AI Challe...
Testing Gemini 3 Pro for Archives
AI Studio: Historical Document Resear...
AI Transcription Blocks and Historical...
Armenian Manuscript Transcription A...
Armenian Translation Forensic Analysis
Historian's Digital Evolution Analysis
Armenian Gospel Text Analysis
Testing LLM Visual Grounding for Hist...

Gemini 3 Pro API Cost Analysis
Visual Grounding for HTR Accuracy
Gemini 3 Pro Cost and Quality Analysis
Historical Document Markup and Tra...
AI Translation Comparison: Armenian ...
Visualizing Armenian Manuscript Coo...
Calculate the in and out token use for...

I want to recognise the lines and regi...
AI's Legal Industry Transformation Re...
Simplexico: AI Legal Training Consult...
Critiquing Legal Tech Training Strategy
AI for Historical Research Workflow
Legal AI Trends and Firm Strategies
Gemini AI Development and Deep Re...
Marinelives.org Site Map Alternatives
Wiki Text Review: Dates and Links
AI Research and Archival Science Dis...
Transkribus Conference Date Correct...

AI Research Assistant JSON Export
GenAI Research Assistant Role Defined

Planning Generative Lives Software P...
CER/WER Analysis: LLM Limitations a...
Download Google Maps Data History
Optimizing Gemini for HTR
Google AI Studio API Key Setup
Historian's AI and Digital History
Handwriting Recognition Cost Estima...
Gemini Pro Account and API Access
Understanding SOTA Reasoning in AI

Building Simulators With Sequential P...
Identifying Presentation Date and Rel...
Identifying Presentation Date and Rel...

Admiralty Deposition Transcription a...
Mariner Behavior Research Deep Dive
Analyzing Historical Maritime Behavior
Analyzing Historical Data: Literacy an...
Mariner Behaviors: Analysis and AI A...
AI Model Version and Capabilities Inq...
Accessing Gemini 3 Pro

Imagine these conversations linked in a Reddit style structure, ideally across platform?

There are already project folders for Gemi, Claude and ChatGPT with various names

I am capturing my experiments in an experiment notebook

Navigation

Search document

Headings Pages Results

- DATE: Sunday, November 30th 2025
 - EXPERIMENTS
 - EXPERIMENT 1
 - EXPERIMENT 2
 - EXPERIMENT 3
 - EXPERIMENT 4
 - EXPERIMENT 5A
 - EXPERIMENT 5B
 - EXPERIMENT 6A
 - EXPERIMENT 6B
 - EXPERIMENTAL ANALYTICS
 - ANALYTICS 1
 - ANALYTICS 2
 - ANALYTICS 3
 - ANALYTICS 4
 - DOCUMENTATION
- DATE: Monday, December 1st 2025
 - JSON SCHEMA DESIGN
 - EXPERIMENTS
 - Experiment 7
 - Experiment 8
 - Experiment 9
 - Experiment 10
 - Experiment 11
 - Experiment 12
- DATE: Tuesday, December 2nd 2035
 - Experiment 13
 - Experiment 14
 - Experiment 15
 - Experiment 16
 - Experiment 17
 - Experiment 18

HTR_Experiment_Notebook_V1.1_09122025

TITLE: HTR Research Notebook
AUTHOR: Colin Greenstreet
VERSION: V1.1
DATE OF ORIGIN: V1: December 1st 2025; V1.1 December 9th 2025
PURPOSE: Document daily research goals, experimental design, and experimental results
PAGES: pp.64

DATE: Sunday, November 30th 2025

EXPERIMENTS

EXPERIMENT 1

Name: Experiment 1: HCA 13/63 f.1r
Design: Temperature = 1.0; Thinking = High; Wang JSON Schema V1.0 [November 30th 2025]
Content: HCA 13/63 f.1r
Intent: Test different combinations of temperature and thinking level
User prompts:

Prompt One:
Transcribe this image by executing you systems instruction

Prompt Two:
Now extract the "transcription" lines from the JSON and list all lines sequentially preserving the line break as plain text

Google AI Studio link:
[https://drive.google.com/file/d/1fRRd3kYEv8j7Vf3yC3VpCmAgAN4JmNoi/view?usp=sharing,](https://drive.google.com/file/d/1fRRd3kYEv8j7Vf3yC3VpCmAgAN4JmNoi/view?usp=sharing)

<https://aistudio.google.com/app/prompts?state=%7B%22ids%22:%5B%2217bfOUixRuEPKQyD8D7vy0tKnunBzs4u6%22%5D,%22action%22:%22open%22,%22userId%22:%2211351180775838411094%22,%22resourceKeys%22:%7B%7D%7D&usp=sharing>

EXPERIMENT 2

Name: Experiment 2: HCA 13/63 f.1r
Design: Temperature = 0.2; Thinking = High; Wang JSON Schema V1.0 [November 30th 2025]
Content: HCA 13/63 f.1r
Intent: Test different combinations of temperature and thinking level
User prompts:

Prompt One:
Transcribe this image by executing you systems instruction

Prompt Two:
Now extract the "transcription" lines from the JSON and list all lines sequentially preserving the line break as plain text

Google AI Studio link:
[https://drive.google.com/file/d/1fRRd3kYEv8j7Vf3yC3VpCmAgAN4JmNoi/view?usp=sharing,](https://drive.google.com/file/d/1fRRd3kYEv8j7Vf3yC3VpCmAgAN4JmNoi/view?usp=sharing)

https://aistudio.google.com/app/prompts?state=%7B%22ids%22:%5B%221zBnZ8XSbeHt0vOqI4lWIKVN_upBpD-%62%5D,%22action%22:%22open%22,%22userId%22:%2211351180775838411094%22,%22resourceKeys%22:%7B%7D%7D&usp=sharing

1

HTR_Experiment_Notebook_V1.1_09122025

EXPERIMENT 3

Name: Experiment 3: HCA 13/63 f.1r
Design: Temperature = 0.2; Thinking = High; Wang JSON Schema V1.1 [November 30th 2025]
Content: HCA 13/63 f.1r
Intent: Test revised and expanded JSON Schema, which builds in learning from comparison of outputs from Experiments One and Two, and introduces concept of parent regions and child lines [See: [Forensic Analysis of Experimental Data: Experiments 1 & 2](#)]

User prompts:

Prompt One:
Transcribe this image by executing you systems instruction

Prompt Two:
Now extract the "transcription" lines from the JSON from the parent regions and child lines in a logical order as plain text preserving the line breaks

Google AI Studio link:
[https://drive.google.com/file/d/1fRRd3kYEv8j7Vf3yC3VpCmAgAN4JmNoi/view?usp=sharing,](https://drive.google.com/file/d/1fRRd3kYEv8j7Vf3yC3VpCmAgAN4JmNoi/view?usp=sharing)

<https://aistudio.google.com/app/prompts?state=%7B%22ids%22:%5B%2217bfOUixRuEPKQyD8D7vy0tKnunBzs4u6%22%5D,%22action%22:%22open%22,%22userId%22:%2211351180775838411094%22,%22resourceKeys%22:%7B%7D%7D&usp=sharing>

EXPERIMENT 4

Name: Experiment 4: HCA 13/63 f.4r
Design: Temperature = 0.2; Thinking = High; Wang JSON Schema V1.2 [November 30th 2025]
Content: HCA 13/63 f.4r
Intent: Test revised and expanded JSON Schema, which builds in learning from analysis of outputs from Experiment Three [See: [Forensic Analysis of Experimental Data: Experiments 1, 2 & 3](#)]

User prompts:

Prompt One:
Transcribe this image by executing you systems instruction

Prompt Two:
Now extract the "transcription" lines from the JSON from the parent regions and child lines in a logical order as plain text preserving the line breaks

Google AI Studio link:
[https://drive.google.com/file/d/1fRRd3kYEv8j7Vf3yC3VpCmAgAN4JmNoi/view?usp=sharing,](https://drive.google.com/file/d/1fRRd3kYEv8j7Vf3yC3VpCmAgAN4JmNoi/view?usp=sharing)

<https://aistudio.google.com/app/prompts?state=%7B%22ids%22:%5B%2217bfOUixRuEPKQyD8D7vy0tKnunBzs4u6%22%5D,%22action%22:%22open%22,%22userId%22:%2211351180775838411094%22,%22resourceKeys%22:%7B%7D%7D&usp=sharing>

EXPERIMENT 5A

Name: Experiment 5A: HCA 13/63 f.4v
Design: Temperature = 0.2; Thinking = High; No Systems Instructions
Content: HCA 13/63 f.4v
Intent: Accidentally ran the intended design for Experiment B without including the planned Systems Instruction for Experiment 5B

User prompts:

Prompt One:

2

I am also using Gemini 3 Pro Preview prompted by me to draft methods papers for different history tech topics

Application of Simulated Nesting Learning for Ottoman Turkish Transcription

Memorandum_Application_of_Simulated_Nested_Learning_for_Ottoman_Turkish_Transcription_V1.0_09
122025

MEMORANDUM

TO: Colin Greenstreet
FROM: Gemini 3 Pro Preview, prompted by Colin Greenstreet
DATE: December 9, 2025
VERSION: V1.0
SUBJECT: Application of Simulated Nested Learning for Ottoman Turkish Transcription

1. Executive Summary

This memorandum outlines the architectural strategy for transcribing **Ottoman Turkish**—a low-resource, high-ambiguity script—using the **Gemini 3 Pro Preview** environment.

We explicitly address the question of methodology: **Yes, this use case utilizes a simulation of "Nested Learning."** The complexity of Ottoman script (missing vowels, ambiguous diacritics, and multi-lingual vocabulary) requires a system that can decouple **Visual Perception** (Fast Context) from **Linguistic Prediction** (Slow Context).

We propose a "**Three-Loop**" Architecture that forces the model to emulate the hierarchical optimization described in the "Nested Learning" literature. This prevents the Language Model (Loop 3) from "hallucinating" plausible but incorrect words when the visual evidence (Loop 1) is unclear.

2. Goal

To achieve high-fidelity transcription of Ottoman Turkish by creating a prompt architecture that mimics **Mult-Level Optimization**. The specific objective is to identify and flag "Visual Ambiguities" (e.g., missing dots) before the semantic engine attempts to resolve them, thereby reducing the error rate in "Generative Smoothing."

3. Methods

3.1 Architecture: The "Three-Loop" Simulation

We cannot alter the physical weights of Gemini 3 Pro. However, we can structure the **Context Window** to simulate three nested optimization loops:

- Loop 1 (Visual Grounding - The "Inner Loop"):**
 - Function:** Raw sensory processing.
 - Constraint:** The model is forbidden from predicting words. It must list visual primitives (*Rasm* strokes, *İjam* dots) and flag pixel-level ambiguities.
- Loop 2 (Context Compression - The "Middle Loop"):**
 - Function:** Orthographic filtering.
 - Constraint:** The model filters the visual data through the rules of Ottoman script connectivity, rejecting candidates that are visually impossible, regardless of meaning.
- Loop 3 (Semantic Execution - The "Outer Loop"):**
 - Function:** Final probabilistic selection.
 - Action:** The model selects the best candidate based on sentence syntax and vocabulary (Arabic/Persian/Turkish roots).

1

Memorandum_Application_of_Simulated_Nested_Learning_for_Ottoman_Turkish_Transcription_V1.0_09
122025

3.2 Adversarial Robustness Testing

To validate that the "Inner Loop" is functioning, we employ **Adversarial Image Generation** using the **Nano Banana Pro** (Gemini 3 Image) model.

- Protocol:** We generate 5 variations of a single source line:
 - Control (Original).
 - No Dots (Ambiguity Test).
 - Added Noise/Ink Blots (Occlusion Test).
- Success Metric:** The system passes if Loop 1 flags the ambiguity in the modified images. If Loop 3 outputs the "correct" word despite the missing dots *without* flagging the error, the Nested Learning simulation has failed (collapsed into hallucination).

3.3 Non-Generative Layout Analysis

For full-page segmentation, we utilize a "**Locator Protocol**" rather than generative image extraction. The model outputs JSON bounding boxes ([ymin, xmin, ymax, xmax]) which are then processed by a Python script (Appendix A.1) to ensure pixel-perfect crops without generative drift.

4. Bibliography (Verifiable Sources)

- Behrouz, A., et al.** (2024/2025). "Nested Learning: The Illusion of Deep Learning Architectures." (Available as Preprint/OpenReview), URL: <https://openreview.net/forum?id=nbMefvNb7A>
- Alayrac, J.B., et al.** (2022). "Flamingo: a Visual Language Model for Few-Shot Learning." NeurIPS, URL: <https://arxiv.org/abs/2204.14198> (Foundational work on interleaving visual/text data).
- Google DeepMind.** (2024). "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context." Technical Report, URL: <https://arxiv.org/abs/2403.05530> (The basis for the long-context visual reasoning used here).

Appendix A: Code Implementation (Ottoman)

A.1 Layout Extraction Script (The "Cookie Cutter") This script takes Gemini's JSON coordinates and crops the high-res image.

```
Python
import json
from PIL import Image
import os

def crop_regions_from_json(image_path, json_path, output_dir):
    img = Image.open(image_path)
    width, height = img.size

    with open(json_path) as f:
        data = json.load(f)
    |
```

2

Taxonomy-based Vision Analysis of HCA 13/ Signoffs

Memorandum_Taxonomy-based_Vision_Analysis_HCA_13_Signoffs_V1.0_09122025

MEMORANDUM

TO: Colin Greenstreet
FROM: Gemini 3 Pro Preview, prompted by Colin Greenstreet
DATE: December 9, 2025
VERSION: V1.0
SUBJECT: Taxonomy-Based Vision Analysis for HCA 13/ Signoffs

1. Executive Summary

This memorandum details a **strategy for analyzing Frontmatter and Signoff blocks in High Court of Admiralty (HCA 13) depositions**.

This HCA task is a **Classification and Extraction** problem. The goal is not to "read" obscure text, but to categorize human motor control (literacy) and extract clear entities.

Therefore, we utilize a "**Structured Vision**" Architecture. This relies on the model's intrinsic multimodal reasoning to link specific pixel regions (the signature) with semantic context (the biography) using a rigid, pre-defined taxonomy.

2. Goal

To create a structured, statistically analyzable dataset of 17th-century literacy. This involves:

- NER Extraction:** Accurately parsing "Front Matter" (Name, Age, Occupation).
- Forensic Classification:** Categorizing the "Signoff" (Signature vs. Mark) using a fixed schema.
- Data Linkage:** Saving the visual evidence (pixel crop) alongside the biographical data in Excel.

3. Methods

3.1 Taxonomy: "Anchor & Compare"

To ensure consistency across 100+ images, we employ a "**Gold Standard**" **Taxonomy** in the System Instructions. The model is forbidden from using free-text classifications. It must select from Enums:

- Class:** ["Full Signature", "Initials", "Mark", "Proxy"]
- Morphology:** ["Simple Cross", "Merchant Mark", "Geometric"]
- Fluency:** ["High", "Medium", "Low"]

This "Anchor" method stabilizes the model's output, ensuring that "Image 1" and "Image 100" are judged by the same criteria.

3.2 Generative Exclusion

Nano Banana (Image Generation) is strictly excluded. Historical forensic analysis requires raw, unaltered pixels. Any generative "cleanup" of a signature would destroy the evidence of literacy (e.g., tremors, ink blots). The workflow uses strictly **discriminative** vision (looking/cropping), not generative vision.

1

Memorandum_Taxonomy-based_Vision_Analysis_HCA_13_Signoffs_V1.0_09122025

3.3 The Excel Pipeline

A Python pipeline (Appendix B.1) is used to automate the workflow:

- Send image to Gemini 3 Pro.
- Receive JSON with biographical data and signature coordinates.
- Crop the signature from the original file.
- Insert the crop into an Excel cell for side-by-side human review.

4. Bibliography (Verifiable Sources)

- Wei, J., et al.** (2022). "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models." NeurIPS, URL: <https://arxiv.org/abs/2201.11903> (The basis for the step-by-step reasoning used in the taxonomy analysis).
- Liu, H., et al.** (2023). "Visual Instruction Tuning." (LLaVA Paper). NeurIPS, URL: <https://arxiv.org/abs/2304.08485> (Demonstrates the capability of VLM architectures to follow complex instruction schemas).
- Google DeepMind.** (2024). "Gemini 1.5 Report." (Specifically the section on Document Understanding and mmuU benchmarks). URL: <https://arxiv.org/abs/2403.05530>.

Appendix A: Code Implementation (HCA)

A.1 Excel Dataset Builder This script automates the creation of the dataset with embedded images.

```
Python
import google.generativeai as genai
import json
import os
from PIL import Image
from openpyxl import Workbook
from openpyxl.drawing.image import Image as ExcellImage

def process_hca_folder(folder_path, output_excel):
    wb = Workbook()
    ws = wb.active
    ws.append(["file", "Name", "Age", "Role", "SignType", "Description", "Signature_Img"])

    row = 2
    for file in os.listdir(folder_path):
        # ... [Gemini API Call using System Instruction B.2] ...
        # Assume 'response' contains the JSON model output

        data = json.loads(response.text)
        bio = data["file_analysis"]["biographical_data"]
        lit = data["file_analysis"]["literacy_event"]

        # Write Text Data
        ws.cell(row=row, column=1, value=file)
        ws.cell(row=row, column=2, value=bio["name"])
        |
```

2

[Discussion]

When is a supportive ai a chatbot, an assistant, or (sounds impressive) an agent?

chatbot

Personas
Systems instructions
User prompts

Discuss
Jan
2026

assistant

ChatGPT-4o
Uploadable files
Instructions
Publishable
Moneytisable

Gemini 3 Pro Preview
App or desktop based
Designed via Google AI Studio
Control of interface
Control of workflow
Publishable
Moneytisable

Discuss
today
[Dec
2025]

agent

Claude Opus 4.5
Google 3 Pro Preview

Discuss
Mar
2025

Agentic browsers
Single autonomous agents
Multiple autonomous agents
Domain specific agents
Function specific agents [customer service; contract analysis]
Agent friendly protocols [e.g. MCP]
Briefing and managing agents

My thinking about Research Assistants has developed since mid-2024

GPTs

- Explore
- Framework Finder
- Knowledge Graph Builder
- Hitchcock
- RootMap
- Aurelius-EM
- Aurelius-Umpire
- Aurelius-Proctor
- Aurelius-Scrivener
- H2O
- Aurelius-NER
- Wallis
- Aurelius-HTR
- Aurelius-Mariner
- Fight-My-HOA
- Aurelius-SIM
- Aurelius-Merchant
- Aurelius-Archives [Updated ...]
- Aurelius-CSV
- Literacy

Hitchcock
By C S Greenstreet

Tim Hitchcock's digital research assistant

Can you help me visualize jury data?
What type of juries existed in C18th England?
How can I model the size of a juror pool?

Aurelius-Archives [Updated Oct 2025]
Last edited Nov 24 ...

Create Configure Preview Model 40

Name: Aurelius-Archives [Updated Oct 2025]

Description: Helps historians interrogate research resources and metadata provided by archives and libraries worldwide

Instructions: Helps historians interrogate research resources and metadata provided by archives and libraries worldwide

Interrogates TNA research guides from internal knowledge:
#Apprentices and masters
#Bankrupts and insolvent debtors

Conversations with your GPT can potentially include part or all of the instructions provided.

Conversation starters:
What research guides do you have available?
Can you write a script to call the TNA-API?
Can you geolocate TNA records?

Knowledge: Conversations with your GPT can potentially reveal part or all of the files uploaded.
Wiki_Base_SPARQL_R_Ver Document

Upload files

Recommended Model: Recommend a model to the user, which should be used by default for best results.
No Recommended Model - Users will use any model they prefer

Capabilities: Web Search Canvas Image Generation Code Interpreter & Data Analysis

+ Ask anything

GL

Generative Lives

Socratic Research Assistant V4.4

ENTER YOUR NAME TO BEGIN

Initialize Research Session

OR

Load Saved Session (.json)

Log in via the welcome screen leads to a short response from the Research Assistant proposing a calibration process to understand the user's risk appetite for inference

The image shows the welcome screen of the Generative Lives Socratic Research Assistant V4.4. At the top center is a red circular logo with the letters 'GL'. Below it is the title 'Generative Lives' in a large, bold, dark blue font. Underneath the title is the subtitle 'Socratic Research Assistant V4.4'. A text input field is labeled 'ENTER YOUR NAME TO BEGIN'. Below the input field is a large dark blue button with the text 'Initialize Research Session' in white. To the right of this button is the text 'OR' in a small blue font. Below 'OR' is another text input field labeled 'Load Saved Session (.json)'. To the right of the main interface is a vertical sidebar containing a message from the Research Assistant. The message reads:

Hello Colin. I'm ready to support your research.
I am designed to operate in two modes:
Pure Fact Hunting: Retrieving clearly cited, verifiable evidence.
Inference Partnering: Helping you connect facts to support or challenge your hypotheses.
To work well together, I need to understand your 'Risk Appetite' for inference. When you tell me during this working session you want me to help with inferences, where on a spectrum do you want me to be? From 'very cautious, only building off inferences made by others in sources' to 'bold, using my native inference skills'?
Please click the '**Calibrate Assistant**' button below to define my operating parameters for this session.

Calibrate Assistant

Capture Analysis & Evidence

Below the sidebar, another section displays the message: 'Calibration Applied: [Mode: CAUTIOUS]'. It continues: 'I have updated my operating parameters based on your custom instructions. What is your first research question?' followed by a 'Capture Analysis & Evidence' button.

Click here for [[demo](#)]

[Discussion]

I want to take our remaining minutes to identify future use cases for assistants and agents which interest members of our collaboratory

Here is a thought starter

AI Agents in Historical Research ▾

I want to find and explore the work of historians and computer scientists talking about in 2024 but especially in 2025 about the potential for LLM based research agents to assist in historical research. Suggest how we can do this together. Answer Socratically.

Show thinking ▾

This is a rapidly evolving area where the conversation has shifted significantly between late 2024 and mid-2025. We are moving from "LLMs as search engines" to "LLMs as autonomous agents" that can plan, query, and verify.

To explore this together effectively, I will guide you through the current landscape using a Socratic approach. I want to help you build a mental map of the field rather than just giving you a list of links.

Phase I: Defining the "Agent" in History

In 2024, most historians used LLMs for summarization or retrieval (RAG). In 2025, the focus is on "Agentic Search"—systems that can autonomously break down a complex historical question, execute multiple searches, write code to analyze data, and self-correct.

Consider this: If an AI can not only find a source but also write Python code to analyze a 19th-century census database and then cross-reference it with a map, is it still just a tool, or has it become a methodological partner?

To start our exploration, which of these two "2025 frontiers" interests you more?

1. **The Archival Agent:** An AI that navigates unstructured text (like OCR'd letters) to find narrative connections.
2. **The Data Agent:** An AI that turns natural language questions into code (SQL/Python) to query structured historical databases (like tax records or censuses).

Phase II: Key Case Studies & Papers (2024-2025)

While you consider that, let's look at where the actual work is happening. I have identified three specific "clusters" of research from late 2024/2025 that we should examine.

1. The "Venice Cadastre" Approach (EPFL / Venice Time Machine)

A seminal paper from late 2025 titled "*LLM Agents for Interactive Exploration of Historical Cadastre Data*" demonstrates an agent that converts natural language into SQL to analyze 200 years of property data in Venice.

- **Why it matters:** It proves agents can handle "dirty" historical data better than strict traditional algorithms.
- **Question for you:** Does your interest lie in quantitative history (*cliometrics*), or are you skeptical of reducing history to data points?

2. The "Deep Research" Paradigm

A trend in computer science (highlighted in August 2025 surveys on "Deep Research Agents") moves beyond single-query answers. These agents perform "iterative retrieval"—they search, read, realize they are missing context, and search again.

- **Why it matters:** This mimics the actual workflow of a historian.
- **Question for you:** Would you trust an agent to determine when it has "enough" context to answer a historical question, or is that a judgment call only a human can make?

3. The "Epistemic Infrastructure" Debate

A 2025 preprint titled "*Large Language Models for History, Philosophy, and Sociology of Science*" argues that LLMs are not neutral tools but "epistemic infrastructures" that encode specific views of history based on their training data.

- **Why it matters:** It challenges the "objectivity" of agent-based research.
- **Question for you:** If an agent is trained primarily on digitized English-language archives, how might it distort a research project on, say, non-Western oral histories?

Phase III: Where to Look & Who to Follow

To stay on the cutting edge of this *right now*, we need to look at specific venues.

• **The Venues:**

- **The Oxford AI Summit (June 2025):** Focused specifically on "Autonomous AI Agents."
- **HKADH 2025 (Hong Kong):** The Digital Humanities conference there has explicitly called for papers on "AI Agents" in cultural analytics.
- **NeurIPS 2025:** Look for the **HiST-LLM** benchmark paper. It showed that while LLMs are great at reasoning, they often fail at "expert-level" historical fact-checking compared to simple graduate student tasks.

Let's synthesize this.

You asked to explore this *together*. To do that, I need to know where you want to drop the anchor.

Based on the above, which path should we take next?

1. **The "Builder" Path:** "I want to see how these agents work technically. Show me the architecture of the Venice project or the 'Deep Research' agents."
2. **The "Critic" Path:** "I want to explore the failures. Show me the HiST-LLM benchmark results and where these agents hallucinate historical facts."
3. **The "Methodologist" Path:** "I want to know how a historian integrates this into their daily workflow. Let's find Christian Henriot's 2025 case study on AI workflows."

Tell me which path appeals to you, and I will pull the specific details for that section.

Over Christmas I want you each to experiment, suing Google AI Studio (or another platform) with building an agent which delivers on your vision. Try grounding your agent with that one overarching instruction. Try adding in sub-instructions. If you get stuck, and you will, get in touch with me and I will help you.

The Socratic Research Assistant (The Partner): *Can an agent act as a 'research partner' that challenges our thesis rather than just retrieving facts?*

Prompt for the Collaboratory:

"Let's write the job description for this Socratic Research Assistant together. If you could hire a PhD student who had read every document in your archive but had zero 'historical intuition,' what is the **one instruction** you would give them to ensure they don't hallucinate?"

- **What constraints?**
- **What behaviours?**
- **What output?**