**Digital Tools for Humanists, Working with AI**

**Summer School 2025 , University of Pisa**

**AI Meets the Archive: Refining Generative Tools for Historical Research**

**26 June 2025**

**Mary Shelley Letters: LLM-Assisted Historical Analysis**

**Tools:** Claude, ChatGPT, DeepSeek; Google Sheets or templates; network graphing tools (Flourish, Kumu); word processor

**Activity Objectives**

* Use LLMs to assist in the analysis of 11 transcribed letters written by Mary Shelley to investigate the roles of LLMs.
* Compare outputs across Claude, ChatGPT, and DeepSeek for accuracy, nuance, and interpretive value.
* Experiment with prompt engineering and evaluate LLM biases.
* Apply outputs (NER, themes, tone) to construct relationship and thematic maps.
* Develop critical skills for AI-assisted research in history and literature.

Before you start if you do not know who Mary Shelley is or her most famous literary work do a quick google search to become familiar with her.

The Letters are available at: <https://www.dropbox.com/scl/fi/6y8a3j35rv8j97h7uk4tg/Digital-Tools-for-Humanists-Activity-One-18thCenturyCookbook.docx?rlkey=til9ysb400hzk3y4lovyrenni&st=un1kqshn&dl=0>

**This Activity is split into several smaller ones. We will work on this activity at several stages during the day. After each part we will return to the group for a discussion of what we learned.**

**Activity Two, PART 1
Summarisation & Interpretation**

**Task:** Each group should chose 2 or 3 letters from the Mary Shelley files.

Browse the letters.

Upload each letter of your chosen 2 or 3 letters.

Run the following prompts in all three LLMs for each of the three letters:

* Prompt 1: "Summarize this letter in under 200 words."
* Prompt 2: "What is Mary Shelley’s motivation and attitude toward the recipient?" (50 words)

**Deliverables:**

* Create a comparison table: accuracy, tone, emotional signals, hallucinations.

**Mini Report-Back #1 :** Each group shares which LLM performed best on one letter and why.

**Activity Two, PART 2**

**Thematic & Sentimental Evolution**

**Upload all 11 letters either by using the file containing all 11 letters or by uploading the letters individually.**

**Initial Prompt:** "List the central themes of these letters."

**Tasks:**

* Prompt for the themes from all 11 letters in a timeline. (you need to generate prompts for this).
* Imagine two questions and prompt the LLMs to provide you with a response. Reflect on the response.
* How accurate do you think the assessment is? Are there any hallucinations?

**Activity Two – Part 3**

**Prompt Engineering & Meta-Comparison**

**Tasks:**

* Select one or two letters (I recommend Letters 1 & 10)
* Individually, students interpret a letter’s tone, purpose, and emotion.
* Run all through Claude, ChatGPT, and DeepSeek.
* Use the following prompts

Prompt A: “Describe the emotional tone of this letter.”

Prompt B: “Is Mary Shelley expressing sadness or hope in this letter?”

Prompt C: “List any emotional language used and what it suggests about her feelings or state of mind.”

**Mini Report-Back #3:**

Which models respond best to which prompt types?

Does the assessment of the LLMs correspond to your assessment (historian) of the letter? Which model produced the most accurate summary.

**Activity Two, PART 4**

**NER Extraction & Relationship Mapping**

**Using all 11 letters.**

**Prompt One :**

You are a Name Entity Recognition (NER) system.

Follow these annotation rules:

- A PERSON includes full name.

- A TITLE includes role and honorifics.

- An ORGANIZATION includes political, governmental, and corporate entities.

- A LOCATION includes any geographic place.

- A DATE includes full or partial temporal expressions.

Use <> for NER Boundaries to mark the beginning and end of the entities

Could you produce an example that would related to a set of other concepts or labels about the same length for tagging letters.

**Prompt Two**: "Extract all named entities: people, places, events, dates."

Prompt X: You will need to generate further prompts based on the directions you decide to follow in terms of questioning.

**Tasks:**

* Validate outputs across 3 models.
* How did the different models perform on the tagging tasks? Overall accurate.
* Build an entity network: frequency, co-occurrence, affective tone.

**Mini Report-Back #2 :** Discussion of lessons learned.

**Activity Two, PART 5**

**Bias & Hallucination Detection Discussion**

**Tasks:**

* Re-examine LLM outputs for modern assumptions, gender biases, or factual errors.
* Note any hallucinations.

**Discussion:** How do these errors shape or distort historical understanding?

# Mary Shelley Letters –Reporting Worksheet

This worksheet is to be used to record your responses and observations for Activities 1 through 4. Use one page per activity. Submit this completed worksheet at the end of the session.

## Activity 1: Summarisation & Interpretation

Letters Analyzed:

Prompts Used:

1. Summarize this letter in under 200 words.

2. What is Mary Shelley’s motivation and attitude toward the recipient?

LLM Responses Comparison Table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Letter # | Claude Summary | ChatGPT Summary | DeepSeek Summary | Interpretive Notes |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Reflection:

Which model produced the most useful or accurate result? Why?

## Activity 2: Thematic & Sentimental Evolution

Prompts Used:

Initial: List the central themes of these letters.

Custom Prompt 1:

Custom Prompt 2:

Themes Identified (per letter or cluster):

Emerging Themes:

Recurring Themes:

Fading Themes:

Reflection:

How well did the models recognize themes? Were there hallucinations?

Model Output Comparison Table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Prompt | Claude | ChatGPT | DeepSeek | Notes |
| Prompt A |  |  |  |  |
| Prompt B |  |  |  |  |
| Prompt C |  |  |  |  |

## Activity 3: Prompt Engineering & Meta-Comparison

Letter(s) Used:

Prompts:

A: Describe the emotional tone of this letter.

B: Is Mary Shelley expressing sadness or hope in this letter?

C: List any emotional language used and what it suggests about her feelings or state of mind.

Model Output Comparison Table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Prompt | Claude | ChatGPT | DeepSeek | Notes |
| Prompt A |  |  |  |  |
| Prompt B |  |  |  |  |
| Prompt C |  |  |  |  |

Reflection:

Which model responded best to which prompt? Why?

## Activity 4: NER Extraction & Relationship Mapping

Prompt 1: Extract all named entities (people, places, dates, events).

Prompt 2: [Custom NER prompt if applicable]

NER Comparison Table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Entity | Type (Person/Place/Etc.) | Claude | ChatGPT | DeepSeek | Correct? Y/N |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Notes on Entity Networks (e.g., central figures, emotional clusters):

Reflection:

How accurate were the models in tagging? What did you learn from mapping relationships?

**APPENDIX 2: Reflection Questions**

* Which model provided the most reliable historical interpretation? Why?
* Where did AI add value, and where did it distort meaning?
* What insights were uniquely human in origin?
* How can this method be scaled or adapted for other archival or literary analysis projects?