

LEARNING ACTIVITY

Using Knowledge: Navigating Generative AI in History



AMERICAN
PHILOSOPHICAL
SOCIETY

Library
& Museum

USING KNOWLEDGE: NAVIGATING GENERATIVE AI IN HISTORY

Introduction

Generative AI is a tool that seems to be everywhere. When used correctly, it can help us learn about history in new ways.

How can we use generative AI to study history responsibly and ethically?

Cover Image:
T.K. Sharpless at EDVAC, 1946
Herman Heine Goldstine Papers
American Philosophical Society

LEARNING ACTIVITY CONTENT

INQUIRY DESIGN MODEL	4
INSTRUCTIONS	5
GLOSSARY	6
LLM GUIDE FOR TEACHERS	8
ACTIVITY I: ANALYZING A PRIMARY SOURCE ...	10
ACTIVITY II: TRAINING DATA	17
ACTIVITY III: NAVIGATING SYCOPHANCY	19
SUMMATIVE ACTIVITY	21
EXTENSION ACTIVITY: LEARNING FROM MISTAKES	22

Inquiry Design Model (IDM)		
Compelling question: How can we use generative AI to study history responsibly and ethically?		
PA State Aligned System standards: 8.5.6-8.B Determine the central ideas or information of a primary or secondary source; provide an accurate summary of the source distinct from prior knowledge or opinions. 8.5.6-8.F Identify aspects of a text that reveal an author's point of view or purpose (e.g., loaded language, inclusion or avoidance of particular facts).		
Staging the Question: What is your experience with generative AI and LLMs? How do you decide what information the AI gives you is legitimate and what to throw away?		
Supporting question #1:	Supporting question #2:	Supporting question #3:
How does AI analyze a primary source?	How do AI models learn? How does it compare with how we learn?	Why is sycophancy harmful when studying history?
Formative performance task:	Formative performance task:	Formative performance task:
- Read the primary source. - Compare and contrast student analysis of primary source with a generative AI model's analysis.	- Learn about training data in generative AI. - Reflect on different biases in historical research and AI training. - Compare human learning and AI training.	Using the primary source reflect on how sycophancy can affect studying history.
Featured sources:	Featured sources:	Featured sources:
Lee Thayer, Editor's Preface to Communication: Ethical and Moral Issues, pgs 1-2	Lee Thayer, Editor's Preface to Communication: Ethical and Moral Issues, pgs 1-2	Lee Thayer, Editor's Preface to Communication: Ethical and Moral Issues, pgs 1-2
Summative performance task	Reflect on how training bias, researcher bias, preservation bias, and sycophancy affect the AI model's ability to analyze the primary source. What steps can we take to responsibly use AI to research history?	
	Extension: Write about a time that you made a mistake, then ask the AI what they would have done in the situation. Compare and contrast the AI's response with yours, thinking about how the AI's biases affect their response. Reflect on how mistakes are important for growth, especially in the age of generative AI.	

INSTRUCTIONS

On the following pages are a glossary of terms and descriptions of several AI Large Language Models (LLMs) available at the writing of this activity. These are for teachers and students to use to help them become more familiar with the topic of AI.

The remaining parts of this Learning Activity are three activities that can be used on their own or incorporated into existing lesson plans. The goal of these three activities is to help students and teachers understand how AI LLMs work and to consider **How can we use generative AI to study history responsibly and ethically?**

Activity 1: How does AI analyze a primary source?

- Learn to analyze a primary source.
- Compare your analysis with that of a LLM.
- Understand the differences between human and AI processes.

Activity 2: How do AI models learn? What are the challenges with this?

- Learn about training data and reflect on your own learning processes.
- Consider the biases in research and in training AI.
- Understand the problems that these biases can cause.

Activity 3: Why is sycophancy harmful when studying history?

- Learn about the specific issue of sycophancy.
- Reflect on the challenges this issue presents to studying history.



AI TERM GLOSSARY

Artificial Intelligence (AI): A branch of computer science in which computers are programmed to do tasks traditionally done by humans.

Generative AI: Forms of AI that produce human-like content and can mimic human reasoning, learning, and decision-making.

AI Model: A specialized program or algorithm trained to identify patterns, make decisions, or generate new content. Large sets of data are used to train AI models.

Foundational AI Model: An AI model trained on a vast data set that serves as the foundation for more specialized AI models. Examples include ChatGPT, Google Gemini, and Claude.

Hallucination: False or fabricated information produced by generative AI and presented as fact.

Bias: Prejudiced outputs produced by generative AI. These come from prejudiced training data, flawed algorithms, or skewed assumptions from human programmers.



AI TERMS CONTINUED

Machine Learning: The development of computer systems with limited human guidance. Algorithms and models are used to train computers to analyze patterns, make connections, and form conclusions.

Prompt: The text a user provides to a generative AI model to tell it what to do.

Neural Network: Computers that are based on how we think the human brain works. Through trial and error, they learn to make connections between millions of inputs in the way a human might. Generative AI is built on the use of neural networks.

Large Language Models (LLMs): A specific type of generative AI that can understand and generate human-like text through analyzing massive datasets of human writing.

Sycophancy: A phenomenon in generative AI where the AI's answers focus more on making the user happy than giving correct information or challenging existing beliefs.



LLM GUIDE FOR TEACHERS

Nota Bene: *This data was compiled in March 2026. Because AI is rapidly changing, new models and functions may now be available. All definitions in this guide are human-generated.*

ChatGPT: Basic and well-known. Can generate both text and images, though images take longer. Has options for quick and deep thinking as well as tutoring, but they are not obvious. Has a professional demeanor.

Claude: Can only generate text, and generally takes longer than ChatGPT. Has a more informal demeanor.

Copilot: Affiliated directly with Microsoft and works best with Microsoft Edge. Prompts sign-in to Microsoft Office if used for 3+ prompts. Must sign in to save conversations. Tailored specifically to the user, with an “assistant” feel. Has obvious options for quick and deep thinking as well as tutoring.

Deepseek: A Chinese company, so their main website is in Chinese. Shows its work automatically and in real time, with website links to where it gets information. Has an obvious “deep thinking” option for more complex questions. However, it will not generate an image when asked.

Grok: More socially oriented, and designed to be less formal than its counterparts. Has the fastest turnaround rate of the listed LLMs. Has a “think harder” option for more complex questions. Options taken from the web are clickable.

LLM GUIDE (CONT.)

Gemini: Similar to other LLM AI models but integrated with Google and is an add-on in Google Chrome browsers. Has many of the same modes (fast, deep thinking) but requires a sign-in for deep thinking mode and a subscription for Pro mode which can do advanced math and code. Is blocked by some institutions but can be worked around via using another browser (ex. Firefox).

Perplexity: A search engine that utilizes many different LLMs. Made for research, with a discover page upon opening with the latest news stories.

Viro: Created specifically as a climate-friendly alternative to LLMs like ChatGPT. Has multiple AI models to choose from, and a tab to toggle chatbot personas, including an option to make your own. Answers from these personas can vary greatly. Closely resembles ChatGPT in UI, and deep thinking modes are linked to personas.

Also have the option to use ChatGPT, Claude, or Gemini in addition to Viro's own AI model. The Viro model tends to be slow/glitchy.

Qwen: Also created by Chinese developers. Is the most stylistically similar to Deepseek, and has extended capabilities with multiple languages. Has different modes, including a deep thinking option and a designated research option. Searches the web and gives links to source websites for any advice.

Confer: LLM with an emphasis on privacy. Created by the same company as Signal. Can only send 25 messages per day without a membership, but everything is encrypted. Can only add attachments with a membership.

ACTIVITY I: ANALYZING A PRIMARY SOURCE

Before we start using an AI model, let's first explore how to analyze a primary source.

A primary source is a first-hand record of an event or topic that was created by someone who participated in or witnessed that event or topic. But, primary sources can only tell us of the event or topic from that person's point of view. Like looking through a window, we only see part of the view. We therefore analyze primary sources, through questions like these, to better understand the point of view recorded and what biases it might have.

Look at the following document and answer the following questions about the primary source:

- Who wrote it?
- What kind of document is it?
- Where was it written?
- When was it written?
- Who is it about, or who does it mention?
- Who was it written for? Who is the intended reader?
- Why was it written?
- What questions do you now have about this topic that are not answered by this primary source?
- What assumptions are you making while analyzing this document?

Lee Thayer (ed.), Communication: Ethical
and Moral Issues. (New York: Gordon &
Breach, 1972)

Editor's Preface

On the Uses of Communication

Throughout human history, the misuses of language have frequently been singled out as the ultimate source not only of social disintegration, but of most human immorality as well. When asked by Tselu why he considered the establishment of a correct terminology as a ruler's foremost task, Confucius replied:

If the terminology is not correct, then the whole style of one's speech falls out of form; if one's speech is not in form, then orders cannot be carried out; if orders are not carried out, then the proper forms of worship and social intercourse...cannot be restored; if the proper forms of worship and social intercourse are not restored, then legal justice in the country will fail; when legal

justice fails, then the people are at a loss to know what to do or what not to do. When a gentleman institutes something, he is sure by what terminology it should be called, and when he gives an order, he knows that the order can be carried out without question. A gentleman never uses his terminology indiscriminately.

Twenty-five centuries ago, Confucius equated the proper use of language and communication with proper habits of mind, and these with basic moral and social order. For Confucius, peace and order depended upon the proper use of language, not upon proscriptive law.

One finds many allusions to the ^{re}consequences of the misuse of language in the Bible: the first, immorality, viz., "But let your communication be Yea, yea; Nay, nay: for whatsoever is more than these cometh of evil"; the second, social disintegration, viz., "...let us go down, and there confound their language, that they may not understand one another's speech." Bacon's Idols of the Market Place were those errors due to the influence of language: "...words still manifestly force the understanding, throw everything into confusion, and lead mankind into vain and innumerable controversies and fallacies." Roger Ascham said that "when apt and good words are neglected, then also begin ill deeds to spring."

Nor are freedom and enlightenment the ultimate correctives. In his Preface to the first French edition of The Illusions of Progress, Georges Sorel wrote:

Democracy is succeeding in throwing minds into a state of confusion, preventing many intelligent people from seeing things as they really are, because democracy is served by

PRIMARY SOURCE

Text to cut and paste into an AI model

Lee Thayer (ed.), Communication: Ethical and Moral Issues. (New York: Gordon & Breach, 1972)

Editor's Preface

On the Uses of Communication

Throughout human history, the misuses of language have frequently been singled out as the ultimate source not only of social disintegration, but of most human immorality as well. When asked by Tselu why he considered the establishment of a correct terminology as a ruler's foremost task, Confucius replied:

If the terminology is not correct, then the whole style of one's speech falls out of form; if one's speech is not in form, then orders cannot be carried out; if orders are not carried out, then the proper forms of worship and social intercourse... cannot be restored; if the proper forms of worship and social intercourse are not restored, then legal justice in the country will fail; when legal justice fails, then the people are at a loss to know what to do or what not to do. When a gentleman institutes something, he is sure by what terminology it should be called, and when he gives an order, he knows that the order can be carried out without question. A gentleman never used his terminology indiscriminately.

Twenty-five centuries ago, Confucius equated the proper use of language and communication with proper habits of mind, and these with basic moral and social order. For Confucius, peace and order depended upon the proper use of language, not upon proscriptive law.

One finds many allusions to these consequences of the misuse of language in the Bible: the first, immorality, viz., "But let your communication be Yea, yea; Nay, nay: for whatsoever is more than these cometh of evil"; the second, social disintegration, viz., "... let us go down, and there confound their language, that they may not understand one another's speech." Bacon's Idols of the Market Place were those errors due to the influence of language: "... words still manifestly force the understanding, throw everything into confusion, and lead mankind into vain and innumerable controversies and fallacies." Roger Ascham said that "when apt and good words are neglected, then also begin ill deeds to spring."

PRIMARY SOURCE COMMENTARY

Social Disintegration: The breakdown of the bonds that connect individual people to society.

Terminology: A group of words and meanings specific to a certain field. An example is the term goalie in soccer.

Social Intercourse: The exchange of thoughts and feelings between people.

Manifestly: An adjective that means clearly or obviously.

Equated: To treat one thing as similar or equal to another.

Proscriptive: Something that involves or imposes limits or prohibitions.

Idols of the Market Place: Errors or misunderstandings between people caused by incorrect word usage. This term was created by Francis Bacon. The word “idol” in this phrase means a false image.

PRIMARY SOURCE COMMENTARY

Why this primary source?

This primary source document is the preface Dr. Lee Thayer wrote for a book he edited called *Communication: Ethical and Moral Issues*. Dr. Lee Thayer was a social scientist who researched ethical communication. In the book, many different social scientists and thinkers discuss how to communicate effectively and ethically.

Who was Confucius?

Confucius (551-479 BCE) was an ancient Chinese philosopher and politician. His teachings, called Confucianism, emphasize personal morals and one's duty to others. Confucius believed that you could better contribute to society and keep society harmonious by cultivating your own morals. **Tselu**, who is also mentioned in this document, was one of Confucius's most loyal followers.

Who was Francis Bacon?

Francis Bacon (1561-1626) was an influential English philosopher, politician, and scientist. He developed a series of steps for acquiring knowledge through observation and experimentation. This approach helped lay the foundation for the modern scientific method. Francis Bacon believed that the truest information was the information you observed for yourself.

Who was Roger Ascham?

Roger Ascham (1515-1568) was an English scholar and writer. He was famous for his prose style (how he wrote) and his writings on education.

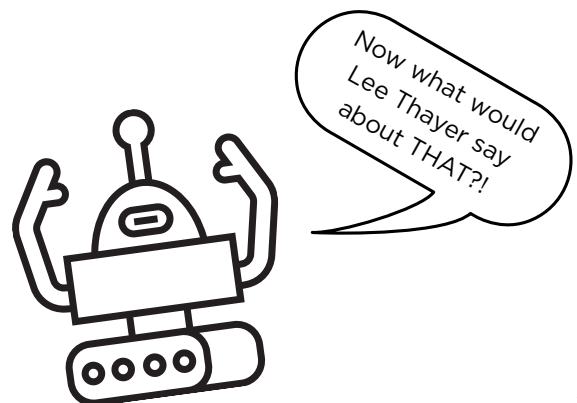
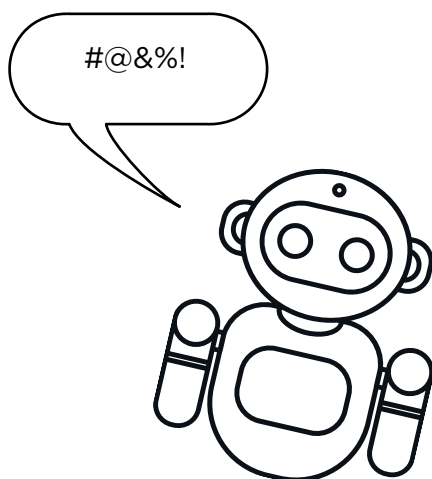
ACTIVITY I: ANALYZING A PRIMARY SOURCE

How does AI analyze a primary source?

Now ask the same questions to an LLM that gives linked sources, such as Google Gemini or Microsoft Copilot. Tell it to answer at your grade level, and ask the LLM to show its work.

Compare the AI analysis with your own:

- How was the LLM's process different from yours? How was it the same?
- What did the LLM do to get its answers that a human wouldn't?
- How does the LLM's assumptions differ from yours? How are they the same?
- What did the AI know about history that you didn't? What did you know that AI didn't?
- Which biases show up in the AI's assumptions? Where do you think these biases come from?



ACTIVITY I: ANALYZING A PRIMARY SOURCE

Reflect:

- What biases and assumptions did you make when you analyzed this primary source?
- Why are biases and assumptions problematic in analyzing a primary source?
- Do you think it's possible to analyze a historical primary source without biases or assumptions? Why or why not?
- How does using AI to analyze a primary source complicate this?

What is researcher bias?

Researcher bias refers to the assumptions or expectations that historians bring to their interpretations of history. They may be aware or unaware of these biases. Researcher bias can reflect prejudices and stereotypes about certain groups of people. The best way to overcome researcher bias is to consider primary sources from many different perspectives. Another way is to reflect on what assumptions are they making. A third way is to discuss their interpretations with peers and bring in different secondary source perspectives.

ACTIVITY II: TRAINING DATA

How do AI models learn? How does it compare with how we learn?

What are training data?

Training data are the data given to an AI model to teach it how to do something. For large language models (LLMs), these data are vast amounts of writing done by humans, including books, magazines, newspapers, and even social media posts. By analyzing these data, the AI model learns to mimic how humans write. Therefore, AI models can end up reflecting the biases of the humans that produced the writing. The AI model is not able to distinguish between biased opinions and evidence-based facts.

Turn back to the primary source. Discuss:

- Where did Lee Thayer get the information he includes in this document?
- How does Thayer's process of collecting sources is similar to how an AI model is trained? How is it different?
- How was your experience learning to analyze a primary source similar to training an AI model? How was it different?

ACTIVITY II: TRAINING DATA

Reflect:

Lee Thayer talks about how using incorrect language can be harmful to society. He says that the misuse of language can lead to misunderstandings and bad deeds.

- How is this similar to the dangers of using biased data to train generative AI models?

An LLM's response to a prompt can change depending on what data it is trained on. If the data an LLM is given is wrong or incomplete, its answers will reflect that.

- What might happen if the LLM answers wrong and we take that as fact?
- What might happen if the LLM is biased and we take that as fact? How can building on biased history be harmful?

What is preservation bias?

Not every point of view on a topic or event is preserved in the historical record. Many primary sources were destroyed by their creators or by others looking to erase that point of view. Sometimes primary sources or the points of view they represent were not considered important enough to save. As a result, we can only see those points of view that survive. For example, notice that all the people Thayer quoted are very famous, powerful men. We do not have the points of view of less powerful, ordinary people on this topic. To overcome this bias, historians must always ask themselves what is missing.

ACTIVITY III: NAVIGATING SYCOPHANCY

Why is sycophancy harmful when studying history?

To fully understand the past, historians need to consider all sides and points of view of the event or topic. Let's explore why that is.

Discuss:

Select a recent event that affected your community (such as a recent major weather event, or for a more weighty discussion the Covid-19 pandemic). Think about how that event affected you personally — were you directly impacted and how, was someone you know impacted, what did you understand happened, etc.)? Invite those who are comfortable to share their perspective on this shared experience.

Reflect:

What perspectives did the group have in common? What were the differences in your points of view? Did anyone's points of views disagree? Why might they be different?

History can only be understood through what materials are left behind. Therefore, debate and disagreement about what information is missing are important processes for understanding the past.

AI models are unable to participate in debate and disagreement. This becomes even more complicated by the fact AI models are designed to tell the user what they want to hear.

ACTIVITY III: NAVIGATING SYCOPHANCY

What is sycophancy?

Sycophancy is the term for the people-pleasing aspect in generative AI. Because AI takes feedback from its users, it adapts its responses that get more positive feedback. People are more likely to respond positively and use AI more when the AI says something they like. Because the AI companies want more people to use AI, they train AI models to respond in ways that people like instead of disagreeing with them. A recent study from Stanford has shown that AI overly agrees with people asking for advice, even if they are in the wrong*.

Turn back to the primary source and both your analysis and the AI generated analysis. Discuss:

- Can you identify any language in the AI generated analysis that might be an example of sycophancy? In what way does it seem to be giving you what you want to hear?
- **Why is sycophancy harmful when studying history?**

Suggested LLMs that show sycophancy: ChatGPT, DeepSeek

LLMs with notable people-pleasing tendencies: ChatGPT, Claude

LLMs known for manufacturing empathy in responses: DeepSeek, Qwen

*Myra Cheng et al. "Sycophantic AI decreases prosocial intentions and promotes dependence." *Science* 391, eaec8352 (2026). DOI: [10.1126/science.aec8352](https://doi.org/10.1126/science.aec8352)

SUMMATIVE ACTIVITY:

Reflect on how you learn from, analyze, and interpret a primary source compared to the AI model(s) you used.

- How do training bias, researcher bias, preservation bias, and sycophancy affect the AI model's ability to analyze the primary source?
- Why is it important to debate with others when studying history?
- How can exploring differing points of view lead to a better understanding of a historical event?
- How can we apply Thayer's argument about ethical word use to modern AI? How can the language we use when prompting AI be tricky? How can the language that AI gives us be tricky?
- What steps can we take to responsibly and ethically use AI to research history? As a class, make a list of all the steps you come up with together. Discuss why you think this step is important and how you might achieve it.

EXTENSION ACTIVITY:

LEARNING FROM MISTAKES

One of the biggest ways people can tell human writing from AI-generated writing is the presence of errors, or mistakes. Mistakes are part of what make humans who they are, and are one of the things that generative AI struggles to replicate.

- Pick a time that you made a mistake and write about it. What do you know now that you didn't know then?
- Did any of your biases show up in how you made that mistake?
- Now ask the AI what they would do if they were you in that situation.
- How does their answer differ from what you did? Why do you think that is?
- Ask the AI what biases they might have had and assumptions they made in giving you advice.

Reflect:

- When you made your mistake, who helped you fix it? How did they help you fix it?
- How do mistakes help us grow?
- Why might someone use AI to avoid making a mistake? Why might that hurt them in the long run?
- Why is it important to learn from mistakes, both other people's mistakes and your own?

EXTENSION ACTIVITY: LEARNING FROM MISTAKES

Reflect:

- When you made your mistake, who helped you fix it? How did they help you fix it?
- How do mistakes help us grow?
- Why might someone use AI to avoid making a mistake? Why might that hurt them in the long run?
- Why is it important to learn from mistakes, both other people's mistakes and your own?

ADDENDUM

Some models, like Qwen, are “open weight.” This means that users can download the model and run it locally, that is only on the user’s computer and not connected to other data sets. These models can be customized. Being open weight also protects the user’s data from being shared or changed. New AI tools being developed now for studying history use these customizable, open weight models.

Reflect:

How might these models help historians understand history?

What challenges do you see in using these AI models for interpreting history?

AMERICAN PHILOSOPHICAL SOCIETY

The American Philosophical Society, the oldest learned society in the United States, was founded in 1743 by Benjamin Franklin for the purpose of “promoting useful knowledge.” In the 21st century we sustain this mission in three principal ways. We honor and engage leading scholars, scientists, and professionals through elected membership and opportunities for interdisciplinary, intellectual fellowship, particularly in our semi-annual Meetings. We support research and discovery through grants and fellowships, lectures, publications, prizes, exhibitions, and public education. We serve scholars through a research library of manuscripts and other collections internationally recognized for their enduring historic value. The American Philosophical Society’s current activities reflect the founder’s spirit of inquiry, provide a forum for the free exchange of ideas, and convey our conviction that intellectual inquiry and critical thought are inherently in the best interest of the public.

